A Publication of the American Psychological Association

Journal of Abnormal and Social Psychology

Founded by Morton Prince, 1906

CONTENTS

Articles

Systematic Changes in Word Association Norms: 1910–1952 3
James J. Jenkins and Wallace A. Russell

Nystagmus: An Integrative Visual Deficit in Schizophrenia 11
Walter W. Zubek

Self-Esteem and Ideal Self-Agreement as Determinants of Interpersonal Perception 19
Stanley Coopersmith

Cooperative Behavior in Dyads as a Function of Self-Appraisal and Verbalization 27
Seymour Rosenberg

Achievement Related Characteristics of Manic-Depressives 37
Joseph Zubin

The Interpretation of Objective Documents as a Function of Socialization and Attitude 45
Melvin Manis

Cognitive Consistency and Attitude Change 57
William J. McGuire

Direct and Indirect Persuasive Effects of Dissonance-Producing Messages 66
William J. McGuire

Continued on inside front cover

Volume 60 MAY 1960 Number

CONTENTS—*Continued*

Some Consequences of Experimental Rejection 359
Albert Pepitone and Chester Wilpizeski
Stimulus Pooling and Social Perception 365
Richard H. Willis
Achievement: Motivation in Fantasy and Behavior 374
Donald M. Braverman, Edward J. Jordan, Jr., and Leslie Phillips
Attitudinal Consequences of Commitment to Unpleasant Behavior 379
Jack W. Brehm
The Frequency of Self-Reference Statements as a Function of Generalized Reinforcement 384
J. Stacy Adams and Browning Hoffman
Interpersonal Perception and Communication in Discussion Groups of Varied Attitudinal Composition 390
Irwin Altman and Elliott McGinnies
Conditioning and Generalization of GSR as a Function of Awareness 396
B. B. Chatterjee and Charles W. Eriksen
The Effects of Subliminal and Supraliminal Suggestion on Verbal Productivity 404
Marvin Zuckerman
Confusion of Figurative and Literal Usages of Words by Schizophrenics and Brain Damaged Patients 412
Loren J. Chapman
Learning a Social Structure 417
Clinton B. De Soto
Repeated Frustrations and Expectations in Hostility Arousal 422
Leonard Berkowitz

CRITIQUE AND NOTES

Frustration, Aggression, and Murder 430
Stuart Palmer
Some Physiological Correlates of Rorschach Body-Image Productions 432
A. D. Davis
Manifest Anxiety and Prestimulus Conductance Levels 437
Gerard J. McDonnell and John A. Carpenter
A Psychological Factor Apparently Associated with Neoplastic Disease 439
Lawrence LeShan and Marcia Reznikoff
Some Relationships between GSR Conditioning and Judgments of Anxiety 440
Harold Gilberstadt and Gene Davenport
Verbal Reinforcement in Schizophrenia 443
Edwin Cohen and Bertram D. Cohen
Ego Disjunction: A Failure to Replicate Trehub's Results 446
Wilbert E. Fordyce and William R. Crow
A Note Concerning Homogeneity of Membership and Group Problem Solving 448
Marvin E. Shaw
The Body-Image Barrier Concept 450
Charles B. Eigenbrode and William C. Shipman

Volume 60 May 1960 Number 3

Journal of Abnormal and Social Psychology

Editor: M. BREWSTER SMITH, *University of California, Berkeley*

Associate Editor: E. J. SHOBEN, JR., *Teachers College, Columbia University*

CONSULTING EDITORS

GORDON W. ALLPORT
Harvard University

URIE BRONFENBRENNER
Cornell University

ROGER BROWN
Massachusetts Institute of Technology

DONALD T. CAMPBELL
Northwestern University

LAUNOR F. CARTER
System Development Corporation

RICHARD CHRISTIE
Columbia University

MORTON DEUTSCH
Bell Telephone Laboratories

I. E. FARBER
State University of Iowa

DONALD W. FISKE
University of Chicago

N. L. GAGE
University of Illinois

ROGER W. HEYNS
University of Michigan

ROBERT R. HOLT
New York University

IRVING L. JANIS
Yale University

HAROLD E. JONES
University of California, Berkeley

HAROLD H. KELLEY
University of Minnesota

LEONARD S. KOGAN
Community Service Society of New York

DANIEL J. LEVINSON
Harvard Medical School

DAVID C. MCCLELLAND
Harvard University

LEO POSTMAN
University of California, Berkeley

ELIOT H. RODNICK
Duke University

NEVITT SANFORD
University of California, Berkeley

SEYMOUR SARASON
Yale University

RICHARD L. SOLOMON
Harvard University

PUBLISHED BY

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, Inc.

The JOURNAL OF ABNORMAL AND SOCIAL PSYCHOLOGY is devoted to both abnormal and social psychology. Its emphasis is on basic research and theory rather than the techniques and arts of practice. Abnormal psychology is broadly defined to include papers contributing to fundamental knowledge of the pathology, dynamics, and development of personality or individual behavior, including deterioration with age and disease. Articles concerned with psychodiagnostic techniques are evaluated with respect to their contribution to an understanding of the psychological principles of diagnosis, those concerned with psychotherapy are judged in terms of their contribution to an understanding of the therapeutic process. Case reports are considered in terms of their heuristic value. Reports which serve to pose or to clarify important theoretical problems, or which promise to be useful in teaching, are often published. From the social area this JOURNAL gives preference to papers contributing to basic knowledge of interpersonal relations, and of group influences on the pathology, dynamics, and development of individual behavior.

The JOURNAL is issued bimonthly in January, March, May, July, September, and November, two volumes per year. The two volumes average 900 pages. The subscription price per year is $16.00; foreign, $16.50. Single issues are $3.00.

Original manuscripts should be sent to the Editor:

M. BREWSTER SMITH
Department of Psychology
University of California
Berkeley 4, California

ARTHUR C. HOFFMAN	HELEN ORR	FRANCES NOYES
Managing Editor	Promotion Manager	Editorial Assistant

Subscriptions, address changes, and business communications should be sent to:

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
1333 Sixteenth Street, N.W.
Washington 6, D. C.

Changes of address must reach the Subscription Office by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

PUBLISHED BY
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, Inc.
Mt. Royal and Guilford Aves., Baltimore 2, Md.
and 1333 Sixteenth Street N.W., Washington 6, D.C.

Second-class postage paid at Baltimore, Maryland, and at additional mailing offices

Copyright 1960 by the American Psychological Association, Inc.

Journal of Abnormal and Social Psychology
Vol. 60, No. 3, 1960

SYSTEMATIC CHANGES IN WORD ASSOCIATION NORMS: 1910–1952[1]

JAMES J. JENKINS
Center for Advanced Study in the Behavioral Sciences
AND WALLACE A. RUSSELL
University of Minnesota

A NORM may be defined as a standard and a standard, as everyone knows, is an unchanging rule to which appeal may be made for precise measurement. Perhaps it is this verbal legerdemain which leads psychologists to overlook the fact that norms which are based on consensus or popularity of responses are especially susceptible to change and require frequent checking in our rapidly changing society.

Word association norms, an example par excellence of the consensus norms, have been commonly assumed to be stable and highly consistent from one time to another. Indeed, investigators rarely raise the issue of the possibility of normative changes but concentrate their attention on differences between groups already known to be disparate. Thus, in addition to the very numerous attempts to compare and contrast clinical groups, many studies have been directed toward comparisons of responses of normal individuals known to differ in some respect. Esper (1918), for example, compared the most frequent responses of American and German students to a standard list of stimulus words and Rosenzweig (1957) contrasted the responses of American and French students; children's responses have been contrasted with adult's responses (Rosanoff & Rosanoff, 1913; Woodrow & Lowell, 1916); the responses of men have been contrasted with those of women (Terman & Miles, 1936; Goodenough, 1942); and the responses of one preprofessional group have been compared with those of another (Foley & MacMillan, 1943).

While these studies have been both interesting and informative, it is illustrative of psychologists' blindness to the normative problems resulting from social change, that little attention has been given to detailed comparisons of even roughly matched groups tested in different years widely separated in time. As far as we have been able to discover only three investigations have mentioned the question.

O'Connor (1934) in the course of an extensive item analysis of the Kent-Rosanoff test data he had collected on a sample of "male factory workers" briefly compared the popular responses to those of the "mixed adult sample" of Kent and Rosanoff (1910). However, he did not pursue the matter further than remarking: "The common response to a stimulus word is, therefore, a scientific reality in the sense that it is rediscoverable by different workers in a new laboratory after a lapse of 15 years," and "the identity of a common response is more reproducible than its frequency of appearance" (O'Connor, 1934, pp. 200–201). Recently, Tresselt and Leeds (1955) compared word association responses collected in 1952 on a small sample, heterogeneous with respect to residence and occupation with "the original norm on the Kent-Rosanoff established in 1927" (actually the 1910 norms). They found a marked increase in primary responses which were responses of opposites (Dark–Light, Black–White) and coexisting pairs (King–Queen, Salt–Pepper). Finally, Dörken (1956) studied 10 high frequency stimulus-response pairs from the Kent-Rosanoff list in various normative collections and discovered that the frequency of these responses seemed to be increasing steadily since 1910.

[1] This study is part of a program of research in verbal behavior being conducted at the University of Minnesota. This study was sponsored by the Office of Naval Research (Contract N 8-onr-66216) under its policy of encouraging basic research. Research assistants who were especially concerned with the material were Evalyn Segal, Walter Mink, and Rochelle Johnson. The Graduate School of the University of Minnesota also assisted with research funds. The senior author is happy to acknowledge his indebtedness to the Social Science Research Council for a Faculty Research Fellowship during the period in which the analysis was conducted.

Apparently, though these writers suggest that changes are underway, a detailed study of changes in word association responses over time remained to be done.

A few years ago the writers fortuitously discovered that the 25-year-old word association norms for students at the University of Minnesota did not adequately represent the associative habits of the present student population. To facilitate the derivation of functional relationships in another problem (Jenkins, Mink, & Russell, 1958) new norms were gathered for the Kent-Rosanoff Word Association Test (K-R) for students in introductory psychology at this university. As these norms were studied and compared with the earlier Minnesota norms collected by Schellenberg (1930), it became obvious that the changes were not minor, trivial adjustments of a few response frequencies (as we had first suspected) but represented a fairly major and systematic change in the norms. Further work suggested that the change was not a peculiarity of sampling or technique but a phenomenon of such scope that a comparison with more diverse populations was indicated. Accordingly, other major American normative collections were studied.

This paper presents first the comparison between the early and later Minnesota norms and then goes on to report the major findings of the comparison of these two sets of norms with the other large-scale normative collections we have been able to discover. The purpose of the study is to attempt to make clear both the magnitude and the nature of the changes revealed by the comparisons of word association norms and to argue that these changes reveal a systematic and important trend in responding to free association tests, which deserves further careful consideration by students of verbal behavior, language and culture.

CHARACTERISTICS OF THE FIVE NORMS

We shall set down at once what is known concerning date of collection, characteristics of samples, and conditions of testing for each of the norms. All five studies employed the same stimulus list and presumably used the same order of presentation.

Norm 1910: Norms collected by Kent and Rosanoff (1910); published in 1910; collection dates not reported.

Testing: Oral form; individual test.

Sample: 1000 "mixed adults."

> Among these subjects were persons of both sexes and of ages ranging from eight years to over eighty years, persons following different occupations, possessing various degrees of mental capacity and education, and living in widely separated localities. Many were from Ireland and some of these had but recently arrived in this country; others were from different parts of Europe, but all were able to speak English with at least fair fluency. Over two hundred of the subjects, including a few university professors and other highly practiced observers, were professional men and women or college students. About five hundred were employed in one or another of the New York State hospitals for the insane, either as nurses and attendants or as workers at various trades, the majority of these were persons of common school education, but the group includes also, on the one hand a considerable number of high school graduates; and on the other hand, a few laborers who were almost or wholly illiterate. Nearly one hundred and fifty of the subjects were boys and girls of high school age, pupils of the Ethical Culture School, New York City. The remaining subjects form a miscellaneous group, consisting largely of clerks and farmers (pp. 38–39).

Norm 1925: Norms collected by O'Connor (1928, 1934); published in 1928; collected in 1925.

Testing: Presumably oral form, individual test; no details are given.

Sample: "1000 adult male factory workers"; no further description of sample.

Norm 1927: Norms collected by Schellenberg (1930); published in 1930, collected in fall, 1927.

Testing: Group test, printed form, written answers; instructions suggest time pressure. Other stimulus words were employed following the K-R list.

Sample: 925 students entering the University of Minnesota. Seventy-nine per cent were freshmen; 21% advanced students. Fifty-seven per cent men; 43% women.

Norm 1933: Norms collected by Keene (1951); published in 1951; collected in academic year 1933–34.

Testing: Oral form; individual test. Other stimulus words were employed following the K-R list.

Sample:

> Five hundred Stanford students, 276 men and 224 women . . . The group included individuals from all four undergraduate levels and also included some graduate students. They were students from the various psychology classes and their friends (p. 43).

Norm 1952: Norms collected by Russell and Jenkins (1954); published in 1954; collected in academic year 1952–53.

Testing: Group test, printed form, written answers.

Instructions suggest a time pressure. (Same instruction with regard to time as Norm 1927.)

Sample: 1,008 students in introductory psychology at the University of Minnesota. Largely sophomore population. About 60% male, 40% female.

COMPARISON OF THE 1927 AND 1952 MINNESOTA NORMS

Since the comparison of old and new Minnesota norms stimulated the remainder of the study and since this is the most readily justified as to comparability of subjects (*Ss*) and method, the Minnesota comparison is given first and in detail. The 1927 and 1952 norms were compared with respect to the following characteristics: (*a*) frequency of primary, secondary, and tertiary responses, (*b*) number of changes in primary, secondary, and tertiary responses, (*c*) relation of the strength of the primary frequencies to the same stimulus words in the two norm sets, the response rank-frequency distribution and the characteristics of the changed response words.

Frequency of Primary, Secondary, and Tertiary Responses

Table 1 gives the average frequency of the primary, secondary, and tertiary responses in each of the norms as well as the total frequency of the three ranks combined. It is clear that a marked increase in the frequency of the most popular response has occurred. The primary response has *on the average* increased in strength by almost 30% of its earlier value; while the most popular response was given by 289 students per thousand in 1927, in 1952 it was given by 375 per thousand. This increase is also manifest at the second ranking response, but the difference has almost disappeared by the third response. Overall, the first three responses in 1952 accounted for the responses of 591 students per thousand, while in 1927 they accounted for only 489 responses per thousand. The change toward more popular responses has obviously been very pronounced.

TABLE 1

AVERAGE FREQUENCY OF THE THREE MOST POPULAR RESPONSES ON THE 1927 AND 1952 MINNESOTA NORMS

Response Rank	Frequency		Percentage	
	1927 (N = 925)	1952 (N = 1008)	1927	1952
1st	267.5	377.5	28.9	37.5
2nd	111.9	137.1	12.1	13.6
3rd	72.8	81.2	7.9	8.1
Total 1st, 2nd, 3rd	452.2	595.8	48.9	59.1

TABLE 2

COMPARISON OF THE RESPONSE WORDS OCCURRING IN THE FIRST THREE RANKS IN THE 1927 AND 1952 MINNESOTA NORMS

1952 Norms	1927 Norms				
	Rank 1	2	3	Lower	Total
Rank 1	71	17	5	7	100
2	14	36	23	27	100
3	8	14	21	57	100
Lower	7	33	51		91
Total	100	100	100	91	

Changes in Popular Response Words

The two sets of norms were compared as to the identity of particular response words. Words ranking first, second, and third in response to a given stimulus in 1927 were compared to the words ranking first, second, and third in 1952. The results are shown in Table 2.

This table shows, as might have been expected, that the words which were primary responses in 1927 tend to be the primary responses in 1952; 71 of the 100 primary responses were identical. Fourteen of the 1927 primaries slipped to the second rank in 1952, eight to the third rank, and seven to ranks below that. The 1927 secondary responses were somewhat unstable; 17 moved up to become 1952 primaries, 14 fell to become tertiaries, and 33 fell to lower ranks; only 36 remained the same. The 1927 tertiaries have an even greater dispersion, as the table shows. Overall, 91 new words entered the top three ranks, and, of course, 91 older words moved out. It is clear that most of this turnover took place at the third rank, with the ranks becoming increasingly more stable as the strength of response increased.

Strength of Primary Responses to Particular Stimuli

In the face of the facts that some primary responses had changed and that primaries in

general had greatly increased in strength, it was of interest to enquire whether the relative strength of a primary response to a given stimulus in 1927 was related to the strength of the primary response to the same stimulus in 1952. Accordingly, three correlations were calculated: (*a*) the correlation between frequency of the primary responses for *all* stimulus words on the two occasions, (*b*) the same correlation for the 71 stimulus-response pairs which had not changed, and (*c*) the correlation for the 29 pairs for which the response had changed. The over-all correlation ($N = 100$ pairs) was + .70; the correlation for same-response pairs was +.65, and the correlation for pairs in which the response had changed was +.48. The three correlations are all significantly different from zero. It is clear that the relative ranks of the strength of primary responses to particular stimuli remain fairly stable over time. Inspection of the scatterplots shows that the primary responses with high strength in the 1927 norms did not tend to change but simply increased in strength as popular responding increased. The primary responses which changed tended to be those which were at relatively low strength in 1927. Such responses were displaced by close competitors in the 1927 norms which moved up to become new low strength primaries.

The Total Response Distribution

The overall distribution of frequencies of responses when plotted against response rank closely approached a Zipf-type distribution. When for the 1952 norms the average frequency of each response rank was plotted against the rank on double logarithm paper, the result was almost a straight line from the first to the 90th rank. After the 90th rank (at which point the average frequency was slightly less than one, indicating that at least some of the response hierarchies had already been exhausted) the curve fell away rapidly toward zero. This distribution may be compared to the rank-frequency distribution for the Kent-Rosanoff 1910 norms reported by Skinner (1937) and the rank-frequency distribution for 200 "other" stimulus words (which were *not* the K-R stimuli) collected by Schellenberg (1930) and plotted by Cook and Skinner (1939). The present distribution is appreciably more linear than either of the earlier distributions and is somewhat steeper, having a slope of approximately −1.4 as contrasted with roughly −1.3 for the Kent-Rosanoff distribution and −1.1 for the Schellenberg "other" words.

The Nature of the Changed S-R Pairs

The stimulus words which evoked differing primary responses in 1952 and 1927 are listed in Table 3 with their old and new responses. The strength of each response is indicated by the percentage of *S*s making that response.

The writers are well aware of the difficulties of classifying associative pairs and are especially wary of "logical" as opposed to "behavioral" classifications. Nevertheless it is clear that at least nine of the 29 stimulus-response pairs in this group that were popular in 1927 were *superordinate* pairs (i.e., Butter-

TABLE 3

STIMULI AND THEIR PRIMARY RESPONSES FOR PAIRS WHICH WERE DIFFERENT IN THE 1927 AND 1952 NORMS

Stimulus	1927 Response	Percentage of Subjects Making the Response	1952 Response	Percentage of Subjects Making the Response
Sickness	Illness	19	Health	37
Mutton	Sheep	34	Lamb	36
Comfort	Ease	13	Chair	12
Short	Long	33	Tall	39
Butterfly	Insect	20	Moth	14
Whistle	Noise	17	Stop	13
Wish	Desire	20	Want	12
Beautiful	Pretty	19	Ugly	21
Window	Glass	26	Door	19
Citizen	Man	21	U.S.	11
Red	Color	23	White	22
Sleep	Rest	25	Bed	24
Working	Labor	11	Hard	13
Earth	Ground	17	Round	13
Trouble	Sorrow	8	Bad	9
Cabbage	Vegetable	40	Head	16
Yellow	Color	29	Blue	15
Justice	Judge	14	Peace	25
Bible	Book	29	God	23
Sheep	Animal	18	Wool	20
Bath	Water	23	Clean	31
Blue	Color	24	Sky	17
Stove	Heat	25	Hot	23
Doctor	Sick	15	Nurse	24
Thief	Robber	22	Steal	28
Lion	Animal	33	Tiger	26
Joy	Happiness	16	Happy	21
Baby	Child	18	Boy	16
Cheese	Food	12	Crackers	11

fly-Insect, Red-Color, Cabbage-Vegetable, Yellow-Color, Bible-Book, Sheep-Animal, Blue-Color, Lion-Animal, and Cheese-Food). If one permits a little ambiguity, the number of pairs classified as superordinate increases (i.e., one may add Whistle-Noise, Wish-Desire, Citizen-Man, etc.) The 1952 responses stand in sharp contrast to this. *None* of the new responses is clearly of a superordinate nature, and only two could be considered to be even marginal superordinates (Sickness-Health? Butterfly-Moth?). It appears from these data that superordinate or classifying responses have decreased in popularity since 1927.

The second major characteristic of the 1927 responses seems to be *synonymity*, but it is difficult to obtain reliable judgments of this relation which are separate from and independent of superordination. The reader may judge for himself how nearly synonymous the stimuli and responses are.

The prevailing characteristics of the 1952 responses seem to be, on the one hand, "coordination" and, on the other hand, "completion" or some form of *sequential relation*. Frequently, these "types" are fused in the new stimulus-response relation. As an example, Cheese–Crackers may be thought of as a pair of coordinates within a class of edibles or as a highly frequent sequential combination in everyday language, "cheese and crackers." The same is true of Lion–Tiger, Doctor–Nurse, Justice–Peace, Red–White, Sickness–Health, etc. Again, systems of classification are inadequate to express the difference between the norms quantitatively.

Overall, the changes in primary responses seem to suggest that current responses are more specific, more concrete, and perhaps more determined by high frequency verbal sequences than the older responses. The older responses which have been discarded tended somewhat more toward abstraction, with the most popular responses being those of superordination and synonymity.

Discussion

At this point the import of the contrast of the 1927 and 1952 norms needs to be stressed. We had suspected from experimental evidence that the word association norms had changed, but we had not been aware of the nature of the change. Before the data were gathered, we had assumed that the 1952 sample would show less agreement and greater diversity in responses than the 1927 sample. The 1952 sample was more heterogeneous (a higher proportion of youth were in school; veterans on the GI Bill represented persons with diverse service backgrounds and in many cases individuals who would not otherwise have been in school, etc.); and the sample was one year advanced in education over the 1927 group (which is supposed to make for more individual responding). In short, we anticipated exactly the opposite findings from those we discovered. Instead of being more heterogeneous, in word association the group was startlingly more homogeneous. There had been marked change, but it was almost all toward conformity.

Clearly there was some stability in free association data across the 25-year period. The high frequency responses tended to persist as the most popular responses, and the ranks of the most popular responses tended to remain the same. On the other hand, we had found a greatly increased tendency on the part of the 1952 *S*s to use the popular responses and a marked tendency to change low frequency primaries of the superordinate sort to other primary responses which tended to be more specific coordinates or completions.

The literature offered little aid in explaining our data. Kent and Rosanoff (1910) had pointed out that their highly educated *S*s made *less* use of the popular response words than the less well educated *S*s in the norm group. Cook and Skinner (1939) had used this finding to explain the distribution of responses on Schellenberg's (1930) 200 "other" stimulus words, where *S*s had in fact used the popular responses less often and the infrequent responses somewhat more often than Kent and Rosanoff's *S*s had for comparable ranks on different stimulus words. But Cook and Skinner were apparently not aware of the fact that on the 100 K-R stimulus words where data were in fact comparable, Schellenberg's sample used the most popular responses *more often* than Kent and Rosanoff's "mixed adults."

We may at this point oppose Kent and Rosanoff's finding and argue that increased education increases the use of popular responses, thus accounting for the differences

between the 1927 and 1910 norms and attributing the difference between the new and the old Minnesota norms to the extra year of education, or we may argue that 1952 sophomores were actually less well educated than 1927 freshmen (and ignore the contrast between the 1927 and 1910 norms) or, finally, we may attribute the difference to something other than educational level such as the methods employed in the various studies or the time-lapse and cultural change between samplings. The writers feel that the last alternative is the only one worthy of serious consideration.

If one were going to attribute the difference to the method employed, one would presumably argue that the group-written form produces more common associates than the individual-oral form, thus accounting for the difference between the 1927 and 1910 norms, and then feel free to introduce some other sort of variable to explain the difference between the 1927 and 1952 Minnesota groups.

Woodworth and Schlosberg (1954), however, feel that the obvious time-pressure of the oral-individual form of the test leads to even greater use of the popular associates than the group-written method. In their chapter on "Association" they compare the results of individual testing of Brown University students with the 1929 Minnesota norms and suggest that the great increase in the frequency of popular responses is a consequence of the difference in method. Inspection of the limited data they present, however, shows a striking correspondence to the 1952 norms, which suggests that the method itself may make only a secondary contribution to the high frequencies observed.

An unpublished study by Clousing (1927) seems to confirm the negligible effect of method (as long as both administrations suggest time pressure). In her study, no differences were found between the group-written and the individual-oral administrations in the frequency of primary, secondary, or tertiary responses. Keene (1951) asserts that there is little difference between the two modes of testing, although he does not present his group data for comparison with his tables of individual data. From the data currently available, explanation of the differences between norms in terms of the differences between individual and group testing does not seem satisfactory.

At this point it was decided provisionally to accept the findings concerning the Minnesota norms as indicating general trends in associative behavior, independent of specific sample or methodological differences. We proposed to test the findings by examining them in the larger context of the 1910, 1925, and 1933 norms. We felt that three hypotheses were testable.

1. Responses of *S*s to the K-R stimuli are increasingly concentrated among the popular responses for all United States adult samples. We proposed to examine the five norms to see if there was evidence for a steady increase in the frequency of the common responses.

2. Words used as responses to the K-R stimuli tend to change slowly over time, with the highest ranking responses having the highest stability. The amount of change for any particular rank should be a function of the time elapsing between the collections of the norm data.

3. "Abstract" responses to stimuli tend to decrease in popularity across the time period encompassed by this study. Specifically, if superordinates are taken as representative of abstractions, they will be found to occur less and less often as responses to stimuli, as the date of the norms becomes more nearly current.

It should be noted that these hypotheses vary in the amount of support they receive directly from the Minnesota comparison already made. The first hypothesis has the most direct support, since the recent Minnesota norms had already been found to be more concentrated in the popular associates than the earlier norms, which in turn were known to be more concentrated than the Kent-Rosanoff. Only if the O'Connor and Keene norms deviated from the trend could the hypothesis be denied. The second hypothesis, however, has many opportunities for confirmation or denial. Each set of norm data may be used as a standard, and the amount of difference between the standard set and other sets may be directly calculated. The third hypothesis similarly has several opportunities for real tests and may be relatively independent of the original data which suggested it (the data on changes in primary responses alone). Here it was decided

to obtain a relatively objective classification scheme, classify all superordinates for the stimulus words, and examine their frequency of occurrence in the several norms. The prediction, of course, was that a decrease in their use would be found over time.

RESULTS OF GENERAL NORM COMPARISON

Frequency of Popular Responses

Table 4 lists the frequences of the most popular responses (first, second, and third rank) for all five norm groups. It is immediately clear that the simple hypothesis of an increase in response frequencies over time will not by itself account for the data. The two norms which are closest together in time, 1925 and 1927, are widely deviate with respect to response frequencies. The 1925 norm is very similar in frequency to the 1933 norm and more like the 1952 norm than it is like the 1910 norm. The 1927 norm is most like the 1910 norm. The similarities and differences here are remarkable in that they cut across the types of samples and methods of administering the tests.

It is tempting to suggest that two factors play a part in the variation found in our data. One factor would be the method of test administration (accepting the Woodworth and Schlosberg [1954] suggestion that oral administration leads to higher popular response frequencies); the other factor would be the general cultural shift which progressively increases the use of common responses. At any one time, one would expect the oral method to yield higher popular response frequencies than the group-written method, but over time both methods would show the increasing drift toward popular responses. The data for primary response frequencies show parallel lines of increase when plotted for oral-individual and group-written administrations separately.

In view of the fact that there is no experimental support for the first factor and that so little is known concerning the actual collection of the 1925 norms, this explanation must be left as a conjecture and as a suggestion and stimulus for further research.

An empirical finding which is quite clear, however, is that at least college populations currently show a very high usage of common responses. The data for Brown University students has already been mentioned as reflecting the high frequencies found in the 1952 and 1933 norms. In addition, support for this assertion is found in the normative compilation of Meals, Herrick, and Merow (1948) at the University of Pennsylvania for 50 of the K-R words. These investigators apparently selected stimuli which elicited high response frequencies as indicated by the 1910 or 1927 norms. Frequency data for the same fifty stimulus words were abstracted from the 1952 norms for comparison with the Pennsylvania data. The proportions responding at each rank are paired for the Pennsylvania and Minnesota norms respectively: first rank, .52, .47; second, .11, .13; and third, .07, .07. It is easily seen that the same intense concentration on popular responses is found for Pennsylvania college students.

Overall, with respect to the first hypothesis, it may be said that the use of popular responses in the free association test is currently at a very high level in college populations, whether they are tested with the individual-oral or group-written method The data strongly suggest that the frequency of popular responses is increasing over time, but that other factors in addition to the general trend are important in determining popular response frequencies.

TABLE 4

AVERAGE FREQUENCY OF THE THREE MOST POPULAR RESPONSES TO 100 STIMULI IN THE FIVE NORMATIVE STUDIES

Response Rank	1910[a]	1925[a]	1927[b]	1533[b]	1952[b]
First	260	338	289	357	375
Second	122	134	121	125	136
Third	77	80	79	78	80
Total 1, 2, 3	459	552	489	560	591

[a] Singular and plural forms of responses were combined in computing frequencies to make these comparable to the other norms.

[b] All frequencies expressed as frequencies per 1000.

Identity of Words Used as Responses

Table 5 lists the percentage of identical responses at each rank level for each pairing of the five normative groups. For the primary responses it is clear that the hypothesis that there is a steady change with time is acceptable. Each norm group shows most overlap with the norm group which is closest chronologically.

The primary responses in the 1952 norms are identical with 73 of the primary responses in the 1933 norms, 71 of the primaries in the 1927 norms, 65 of the primaries in the 1925 norms, and only 58 of the primaries in the 1910 norms. Conversely, the 1910 norms show identical primaries with 81 responses in the 1925 norms, 79 responses in the 1927 norms, 74 responses in the 1933 norms, and 58 responses in the 1952 norms. Over the entire table in similar fashion all the values are in accord with the general hypothesis.

For the secondary and tertiary responses, similar results are found with good conformance to the hypothesis except that the 1925 and 1910 norms are most alike in both cases and the 1933 norms are slightly less similar to the 1927 norms than would have been expected. From either extreme, 1910 or 1952, the progression is quite orderly, however.

While the 1948 Pennsylvania norms referred to above are probably somewhat biased by the selection of stimuli known to give high frequency response, it is interesting to note in passing that they show very high agreement with the 1952 norms. Eighty-two per cent of the primary responses are identical, 42% of the secondary responses are identical, and 32% of the tertiary responses are identical. Inspection of Table 5 reveals that these approximate very closely the percentages of overlap of the 1925 and 1927 norms (83, 48, and 32%) which are separated by a similar short time period.

TABLE 5

PROPORTION OF IDENTICAL RESPONSES IN THE FIVE NORMATIVE STUDIES

Norm	Primary Responses				
	1910	1925	1927	1933	1952
1910	—	81	79	74	58
1925	81	—	83	79	65
1927	79	83	—	80	71
1933	74	79	80	—	73
1952	58	65	71	73	—

Norm	Secondary Responses				
	1910	1925	1927	1933	1952
1910	—	53	44	26	26
1925	53	—	48	47	34
1927	44	48	—	43	36
1933	26	47	43	—	46
1952	26	34	36	46	—

Norm	Tertiary Responses				
	1910	1925	1927	1933	1952
1910	—	36	20	19	10
1925	36	—	32	29	19
1927	20	32	—	27	21
1933	19	29	27	—	29
1952	10	19	21	29	—

TABLE 6

NUMBER OF WORDS DIFFERENT IN THE FIRST THREE RESPONSES (300 words)

Norm	1910	1925	1927	1933	1952
1910	—	69	79	94	123
1925	69	—	66	71	95
1927	79	66	—	88	91
1933	94	71	88	—	77
1952	123	95	91	77	—

Table 6 presents the other side of the consistency picture, showing how many *different* words are found in the first 300 response words for each of the norm pairs. These data conform closely to the hypothesis except in the degree of difference between the 1933 and 1927 norms. In all other respects the table is in accord with the hypothesis.

Considering all the findings, it seems clear that a major determinant of response-overlap or response-identity in word association norms is the time interval between collections of data. Where deviations from this simple prediction pattern do appear, there is a suggestion that the norms derived by the oral-individual method are slightly more alike than would be expected from time lapse prediction. That other variables are probably active is of course conceded, but it is stressed here that time lapse alone accounts for most of the systematic variation.

Use of Superordinates

In order to obtain an unbiased and relatively objective standard for deciding which responses were superordinates, a written test was devised. The test consisted of a set of 100 sentences of the form "—— is a member of the class of ——." Each sentence began with one of the Kent-Rosanoff stimulus words. This

TABLE 7
WORDS DEFINED AS SUPERORDINATES BY THE CRITERION OF THE SUPERORDINATE TEST

Stimulus	Superordinate Response	Frequency (N = 29)	Stimulus	Superordinate Response	Frequency (N = 29)
Red	Color	29	Swift	Speed	21
Yellow	Color	29	Spider	Insect	21
Bread	Food	29	Bed	Furniture	20
White	Color	29	Ocean	Water	20
Green	Color	29	Slow	Speed	20
Butter	Food	28	Scissors	Tool	19
Blue	Color	28	Butterfly	Insect	19
Black	Color	28	Heavy	Weight	18
Lion	Animal	27	Bible	Book	18
Sour	Taste	27	Cabbage	Vegetable	18
Sheep	Animal	26	Blossom	Flower	17
Cheese	Food	25	Sickness	Health	17
Hammer	Tool	25	Music	Art	17
Eagle	Bird	25	Quiet	Sound	16
Chair	Furniture	25	Lamp	Furniture	16
Table	Furniture	25	River	Water	16
Bitter	Taste	23	Salt	Food	15
Sweet	Taste	23	Stomach	Organ	15
Fruit	Food	22	Anger	Emotion	15
Mutton	Meat	22			

test was given to 29 students in an introductory psychology laboratory course. A superordinate response was defined as any sentence completion that was given by 15 or more of the students taking the test. We realize that this procedure is arbitrary in the extreme, but it includes the stimulus-response pairs which most judges would agree are "really" superordinates, provides an unequivocal criterion for counting responses, is unbiased with respect to the normative data to be judged, and excludes words for which superordinates are, to say the least, difficult to define. This procedure yielded agreement as to superordinate for 39 of the 100 K-R stimuli. These stimuli and their superordinates are listed in Table 7 with the frequency with which the superordinate was given.

The occurrences of the superordinate responses to appropriate stimuli were counted in the 1910, 1925, and 1952 norms where complete data were available. The 1927 and 1933 norms could not be used in this comparison because only the three most popular responses are available for each stimulus. Accordingly, a separate tabulation was made of the 26 superordinate responses which were present in the 1927 norms and the 18 superordinate responses which were present in the 1933 norms. The data are presented in Table 8.

Inspection of the table verifies the hypothesis concerning the decline of superordinate responses. The frequency of occurrence of the 39 superordinate responses identified has declined in the 1952 norms to just slightly more than half the frequency manifested in the 1910 norms. The frequency of occurrence of the 26 superordinates present in the 1927 norms (1st, 2nd, and 3rd ranks), even though biased upward for that norm set because of selection, shows a marked decrease from 1910 to 1952. The frequency of occurrence of the 18 superordinates present in the 1933 norms does not even appreciably reflect the selective bias but falls steadily from 1910 to 1952.

A detailed response-by-response comparison confirms this trend as a general case not attributable to a marked change in a few responses. Comparison of the 1910 and 1952 norms shows decreasing frequencies for 33 of the 39 superordinate responses. Further, for what might be called the clearest cases of superordination (those identified on the superordinate test with a frequency of 20 or greater), 24 out of 25 responses have declined. Comparison of the 1925 and 1952 norms shows a decline in 31 of the 39 superordinate responses and 23 of the 25 "strong" superordinate responses. The comparison of the 1910 and 1925 responses, while not as overwhelming, is still conclusive; 30 of the 39 superordinates declined in frequency and 21 of the 25 "strong" superordinates declined in frequency. The

TABLE 8
USE OF SUPERORDINATES AS RESPONSES

	1910	1925	1927	1933	1952
39 Superordinates					
Average percentage	17.5	15.4	—	—	9.6
26 Superordinates[a]					
Average percentage	24.7	22.3	22.6	—	13.9
18 Superordinates[a]					
Average percentage	31.1	28.3	27.9	22.4	19.2

[a] For the 1927 and 1933 norms only the first three responses are available. Twenty-six superordinates occurred in the first three ranks in 1927 and 18 superordinates occurred in the first three ranks in 1933. All those appearing in 1933 also appeared in 1927. The second and third rows of the table are biased upward for 1927 and 1933, therefore.

decline of popularity of the superordinate response appears to be strikingly demonstrated.

Discussion

The first hypothesis concerning the general increase in the frequency of popular responses appears only partially confirmed. It is clear from the data that popular responses are currently at very high frequencies in college populations, but the five normative summaries do not support the notion of a simple trend in this direction over time on all United States adult samples. The 1925 sample shows a much greater use of popular responses than the 1927 sample. The most parsimonious explanation at this time seems to be a two-factor explanation recognizing both the drift toward popular responses over the time-span included in this study and a tendency for the oral-individual test technique to generate more popular responses than the group-written technique.

The second hypothesis concerning the change in identity of popular response words over time seems to be adequately confirmed. Norms collected closest in time show the maximum amount of overlapping in responses (regardless of sample differences), and high frequency responses prove to be much more stable than lower frequency responses. There is a further suggestion that test technique contributes to similarity of responses.

The third hypothesis concerning the decline in the use of superordinates as word association responses is likewise adequately confirmed. A clear and highly significant trend was discovered and the evidence further indicates that the more clearly the responses were identified as superordinates, the more marked the trend was seen to be.

Interpretations of these findings are at best conjectural. The writers have been tempted to discuss the study in terms of the growth of "mass culture," "other-directedness" and similar constructs much used today in characterizing our culture. The enormous growth of the mass media, the standardization of school courses, procedures and text books, and the presumably greater homogeneity of our verbal culture are appealing variables to invoke to explain the phenomena manifest in our data, but no *evidence* can now be marshalled to support such a discussion.

A much simpler explanation for much of the data suggests itself; namely, that test-taking attitudes have been changing over the time period discussed here. The work of Dunn, Bliss, and Siipola (1958), Siipola, Walker, and Kolb (1955), and to a lesser extent that of Flavell, Draguns, Feinberg, and Budin (1958) clearly establishes the importance of the time-pressure set in producing popular and "superficial" responses, such as contrasts and coordinates, and decreasing responses which have a synonymous or superordinate relation to the stimulus. These modern studies of test-taking attitude mirror the effects achieved by Wells in a much older study of practice effects in free association. Wells (1911) gave his *S*s 20 association lists of 50 words (one list per day) and then retested them on the first two lists. He observed a marked decrease in reaction time and a change to "superficial" associations. He reports the qualitative changes in response for each of his six *S*s, and for each one it is remarked that the frequency of superordinate responses decreased (the decreases vary from one-half to two-thirds of the original number of responses so classified).

It would appear that practice effects ("test-wiseness") and time pressure ("test-attitude") bring about much the same effects: so-called superficiality of response, decrease in superordinates, and increased popular responding. If we assume that the oral-individual form maximizes time pressure and that more recent dates give us more and more test-sophisticated *S*s we would seem to have all the necessary variables required to account both for the frequency data and the superordinate data.

It is true of course that "test-wiseness," or "test-docility" of *S*s is after all a general product of our culture and that this explanation may still be related to general questions of cultural change on a grand scale. However, this explanation has the virtue of greater specificity and seems to furnish a point of attack for further investigations which the more sweeping explanation fails to furnish.

It should be noted beyond this, however, that regardless of the explanation given for the increasing frequencies of popular responses and decreasing frequencies of abstract responses, the systematic change in the overlapping of specific stimulus-response pairs as a

function of time-lapse between normative studies is in the main separable from these phenomena. This systematic change appears to the writers to be attributable not only to the drift away from abstract responses but also to changes in the "meaning" of words (e.g., "lamp" in 1910 apparently meant an "oil lamp" in contrast to its current meaning) and to particular recent experiences concerning the words which were not duplicated during the other norm-collection periods (e.g., in 1952, a strenuous presidential campaign was in progress and "whistle-stop" became a very popular associative pair). These changes could profitably be studied as part of a general study of the change in meaning of common words over time.

SUMMARY

A comparison was made of word association norms collected in 1927 and in 1952 for University of Minnesota students. It was found that primary responses had greatly increased in frequency and that the first three responses to the stimulus words accounted for 59% of all responses in 1952 as contrasted with 49% in 1927. Seventy-one per cent of the primary responses were identical. Primary responses were found to be more stable than secondary or tertiary responses. The frequency of the primary response to individual stimuli correlated +.70 between the two norms. Primary responses which had changed, tended to be low-strength superordinate responses which were replaced by coordinates and completions.

Three hypotheses arising from the 1927–1952 comparison were examined with respect to five major collections of free association data: the 1910 Kent-Rosanoff norms, the 1925 O'Connor norms, the 1933 Keene norms, and the 1929 and 1952 Minnesota norms. The first hypothesis, that there is a general tendency for the frequency of popular responses to increase with time, received partial confirmation. It was concluded, however, that other factors, such as testing method probably exercised some effect in addition to the time change.

The second hypothesis was that the words used as responses to stimuli tend to change slowly but systematically over time with the highest ranking responses having the highest stability. This hypothesis was confirmed.

The third hypothesis was that abstract responses (specifically superordinates) have tended to decrease in popularity across the time period of this study. Superordinates were identified for the Kent-Rosanoff stimuli by means of a sentence completion technique. The hypothesis was confirmed.

Possible explanations for these phenomena were mentioned and it was suggested that a general change in test-taking attitudes might account both for the increase in popular responses and the decrease in abstract responses. It was further suggested that the change in individual response terms (aside from the frequency and abstraction changes) might be attributable to changes in the meaning of particular stimuli over time.

REFERENCES

CLOUSING, JOSEPHINE. *The relation between personality trait ratings and free association responses.* Unpublished Master's thesis, Univer. of Minnesota, 1927.

COOK, S. W., & SKINNER, B. F. Some factors influencing the distribution of associated words. *Psychol. Rec.*, 1939, **3**, 178–184.

DÖRKEN, H., JR. Frequency of common associations. *Psychol. Rep.*, 1956, **2**, 407–408.

DUNN, SANDRA, BLISS, JOAN, & SIIPOLA, ELSA. Effects of impulsivity, introversion, and individual values upon association under free conditions. *J. Pers.*, 1958, **26**, 61–76.

EAPER, E. H. A contribution to the experimental study of analogy. *Psychol. Rev.*, 1918, **25**, 468–487.

FLAVELL, J. H., DRAGUNS, J., FEINBURG, L. D., & BUDIN, W. A microgenetic approach to word association. *J. abnorm. soc. Psychol.*, 1958, **57**, 1–7.

FOLEY, J. P., JR., & MACMILLAN, Z. L. Mediated generalization and the interpretation of verbal behavior: V "Free association" as related to differences in professional training. *J. exp. Psychol.*, 1943, **33**, 299–310.

GOODENOUGH, FLORENCE L. The use of free association in the objective measurement of personality. In Q. McNemar & M. A. Merrill (Eds.), *Studies in personality.* New York: McGraw-Hill, 1942. Pp. 87–103.

JENKINS, J. J., MINK, W. D., & RUSSELL, W. A. Associative clustering as a function of verbal association strength. *Psychol. Rep.*, 1958, **4**, 127–136.

KEENE, C. M. Commonality of response on a word-association test: A study of standardization procedures and an attempt to forecast moderate emotional maladjustment. Unpublished doctoral dissertation, Stanford Univer., 1951.

KENT, GRACE H., & ROSANOFF, A. J. A study of association in insanity. *Amer. J. Insanity*, 1910, **67**, 37–96, 317–390.

MEALS, D. W., HERRICK, C. J., & MEROW, ROSALIND.

Scoring sheet for reaction word experiment. Univer. Pennsylvania, 1948. (Mimeo.)

O'CONNOR, J. *Born that Way,* Baltimore: Williams & Wilkins, 1928.

O'CONNOR, J. *Psychometrics,* Cambridge: Harvard Univer. Press, 1934.

ROSANOFF, ISABEL R., & ROSANOFF, A. J. A study of association in children. *Psychol. Rev.*, 1913, **20**, 43–89.

ROSENZWEIG, M. R. Etudes sur les associations des mots. *Année Psychol.*, 1957, No. 1, 23–32.

RUSSELL, W. A., & JENKINS, J. J. The complete Minnesota norms for responses to 100 words from the Kent-Rosanoff word association test. USN Tech. Rep. No. 11, 1954, ONR N8 onr-66216, Univer. Minnesota, 1954.

SCHELLENBERG, P. E. A group free-association test for college students. Unpublished doctoral dissertation, Univer. of Minnesota, 1930.

SHPOLA, ELSA, WALKER, W. NANNETTE, & KOLB, DOROTHY. Task attitudes in word association, projective, and nonprojective, *J. Pers.*, 1955, **23**, 441–459.

SKINNER, B. F. The distribution of associated words. *Psychol. Rec.*, 1937, **1**, 71–76.

TERMAN, L. M., & MILES, CATHERINE C. *Sex and personality: Studies in masculinity and femininity,* New York: McGraw-Hill, 1936.

TRESSELT, M. E., & LEEDS, D. S. The Kent-Rosanoff word association: I. New frequencies for ages eighteen–twenty-one and a comparison with Kent-Rosanoff frequencies, *J. genet. Psychol.*, 1955, **87**, 145–148.

WELLS, F. L. Practice effects in free association. *Amer. J. Psychol.*, 1911, **22**, 1–13.

WOODROW, H., & LOWELL, F. Children's association frequency tables. *Psychol. Monogr.*, 1916, **22** (5, Whole No. 97).

WOODWORTH, R. S., & SCHLOSBERG, H. *Experimental Psychology.* (rev. ed.) New York: Holt, 1954.

(Received July 21, 1959)

Journal of Abnormal and Social Psychology
Vol. 60, No. 3, 1960

NYSTAGMUS:

AN INTEGRATIVE NEURAL DEFICIT IN SCHIZOPHRENIA[1]

WALTER W. LEACH
Mississippi Southern College

THE significance of body image in psychopathology calls attention to proprioceptor mechanisms. One of the most important is the vestibular apparatus because of its reflexive adjustment of the body's position. Through integration of impulses from the semicircular canals at cerebellar and oculomotor centers, an individual maintains equilibirum and orientation to his environment. Mowrer (1932) and Holsopple (1929) demonstrated that semicircular canal eye reflexes maintain normal visual perception during head movements. The observation that some schizophrenic patients have a marked disturbance in body image and orientation to their environment suggests disturbances in their equilibratory processes.

The semicircular canals regulate muscle tonus throughout the body. Angyal (1935, 1936, 1937) describes a schizophrenic syndrome in which a bizarre experience of the body occurs, e.g., an impression of heaviness or lightness. Angyl and Blackman (1940) used the Barany test to compare the postrotation nystagmus of 58 schizophrenic and 20 normal subjects (*S*s). They lacked a recording device and could set no criteria for the eye movements that constitute a nystagmic beat. In addition, they made no observations during stimulation. Objections can be raised concerning the nature of the stimulation in the Barany test (10 rev. in 20 sec. with a sudden stop). This extremely intense and erratic stimulus is greatly beyond appropriate physiological limits and frequently produces distressful side-effects.

The primary purpose of this paper is to compare experimentally groups of schizophrenic and normal *S*s by measuring characteristics of their nystagmus. Amplification of the finding of Angyal and Blackman (1941) required apparatus to provide controlled uniform angular acceleration and to record nystagmus during as well as after rotation. A neural correlate of schizophrenia could provide an objective basis for diagnostic classification. In its broader implications, this experiment is aimed at a unified approach to biological and psychological functioning.

METHOD[2]

Subjects. Seventy-five male inpatients, diagnosed as schizophrenic by the staff of the Lyons, New Jersey, VA Hospital, made up the experimental group. Their ages ranged from 19 to 49 (mean, 29; *SD*, 8). Patients with a record of neurological disease were excluded. Average duration of hospitalization was less than six months.

Ninety-six noninstitutionalized males, ranging from 14 to 76 years of age (mean 26; *SD*, 10), were in the normal control group. Nearly all were university students. The same basis for exclusion applied as for the patient group.

Apparatus. A rotating chair designed by J. Q. Holsopple and the writer provided controlled intensities of uniform angular acceleration. Since a falling weight accelerates at a uniform rate if meeting constant resistance, an essential requirement was that the chair be precisely balanced. During the fall of a weight attached to the axis by a cable, the chair accelerated until the cable unwound. It then decelerated at a uniform rate as the momentum lifted the weight. The total weight of the chair while rotating *S*s of varying weights was kept constant. Each *S* was weighed, and lead ingots in a pan attached to the chair were added or subtracted accordingly.

The intensity of acceleration was determined by the amount of weight attached to the cable. Uniformity and intensity were empirically calibrated at each value of the stimulus. The importance of known uniformity is seen if one considers how unsatisfactory an average value would be in other sense departments.[3]

A standard ophthalmograph (American Optical Company, Model 1520) mounted on the foot plate of the chair was used to record nystagmus during and after stimulation. Start and end of stimulation were signaled directly on the ophthalmographic film. The head rest aligned the vertical axis of the head with that of the chair to stimulate the horizontal semicircular canals.

[1] This paper is based on a dissertation presented to the faculty of Princeton University in partial fulfillment of the requirements for the degree of doctor of philosophy. The writer wishes to express his appreciation for assistance to E. G. Wever and the late J. Q. Holsopple.

[2] For a more detailed description see Leach, 1953.

[3] For a discussion of the problems involved cf., Holsopple, 1929; Mowrer, 1932; and ter Braak, 1936.

Procedure. After *S* was seated in the chair and the head rest was adjusted to prevent movement, he was instructed to look at an "X" a short distance from his eyes. Corneal reflections from each eye were focused onto the film. A record of eye steadiness was obtained before each rotation. *S*s were stimulated in three separate rotations with mild, moderate, and strong accelerations administered in a sequence of 2, 5, and 12° per sec. per sec. Duration of the rotation was 30 sec. at 2°, 21 sec. at 5°, and 16 sec. at 12°. The first and third rotations were counter clockwise; the second was clockwise.

A brief explanation of the purpose was given to each *S*. Cooperation of psychotic *S*s was facilitated by eliminating all cues and any need for communication. The lights of the ophthalmograph were shielded so that none of the surroundings were visible.

Measures. The following measures were made from the nystagmograms at each stimulus intensity:

1. Acceleration latency: time in seconds from beginning of rotation until the completion of the first nystagmus.
2. Deceleration latency: time in seconds from beginning of rotation until the completion of the first reversed nystagmus.
3. Number of acceleration nystagmi during rotation.
4. Number of deceleration nystagmi during rotation.
5. Number of postrotation nystagmi.
6. Duration of postrotation nystagmi: time in sec. until (*a*) no nystagmus appeared for three sec. or more, or (*b*) until response rate dropped by one-half or more. Rotation and postrotation rates were obtained by dividing the number of nystagmi for each period by the duration of each in sec.

The nystagmograms were read by passing the 35 mm. film over a translucent glass in front of a 25-watt light bulb. Since the film was exposed at the rate of ½ inch per sec., time was measured in units of .12 sec. along the edge of a ruler calibrated in sixteenths of an inch. A nystagmic beat was counted for each sequence of slow deviation of the eye tracings from the fixation point and a compensatory saccadic return. The direction of the eye movements was to either the right or left, depending upon (*a*) whether the stimulus was acceleration or deceleration, and (*b*) the direction of rotation. In a clockwise rotation, the slow phase during acceleration was to the left and the compensatory phase to the right. This sequence reversed to right-left during deceleration. In counter-clockwise rotation, these directions were exchanged.

RESULTS

The schizophrenic group showed diminished nystagmic reactivity for all of the measures except deceleration latency. Their deficits during stimulation appeared in their longer reaction times and slower rates of response. Postrotation nystagmus showed a similar reduction. The percentage deficit and the statistical significance of the differences between the schizophrenic patient and the normal control groups are given in Tables 1 and 2.

The amount of deficit varied, depending upon the intensity of stimulation. It was least to strong stimulation and greatest for weak stimulation for all measures except number of deceleration nystagmi. As the stimulus inten-

TABLE 1

COMPARISON OF ROTATION NYSTAGMUS AT 3 STIMULUS INTENSITIES IN 96 NORMAL AND 75 SCHIZOPHRENIC *S*s

	Normal		Patient		Probability	% Deficit
	(*M*)	(*SD*)	(*M*)	(*SD*)		
Acceleration latency in seconds						
2° per sec. per sec.	1.3	1.0	3.0	2.8	0.0001	131
5° per sec. per sec.	1.6	1.4	2.1	1.7	0.05	31
12° per sec. per sec.	1.3	1.0	1.7	1.5	0.05	31
No. of acceleration nystagmi						
2° per sec. per sec.	21.6	8.5	11.6	6.8	0.0001	46
5° per sec. per sec.	16.0	6.7	12.2	6.8	0.0002	24
12° per sec. per sec.	19.1	7.0	14.9	7.2	0.0001	22
Deceleration latency in seconds						
2° per sec. per sec.	23.9	3.3	23.0	3.9	0.152	——
5° per sec. per sec.	17.0	2.2	17.2	3.0	0.63	——
12° per sec. per sec.	13.4	1.3	13.8	1.7	0.10	——
No. of deceleration nystagmi						
2° per sec. per sec.	9.0	1.9	6.2	4.0	0.0001	31
5° per sec. per sec.	6.0	3.1	3.9	2.8	0.0001	22
12° per sec. per sec.	4.1	2.3	2.9	1.9	0.0001	34
Rotation rate: nystagmi per second						
2° per sec. per sec.	0.98	0.33	0.60	0.28	0.0001	39
5° per sec. per sec.	1.06	0.36	0.81	0.41	0.0001	23
12° per sec. per sec.	1.47	0.46	1.16	0.50	0.0001	21

TABLE 2
COMPARISON OF POSTROTATION NYSTAGMUS AT 3 STIMULUS INTENSITIES IN 96 NORMAL AND 75 SCHIZOPHRENIC *Ss*

	Normal		Patient		Probability	% Deficit
	(*M*)	(*SD*)	(*M*)	(*SD*)		
Number						
2° per sec. per sec.	12.7	7.0	6.1	5.4	0.0001	52
5° per sec. per sec.	14.5	8.2	7.4	5.6	0.0001	49
12° per sec. per sec.	13.8	7.8	8.7	6.2	0.0001	37
Duration in seconds						
2° per sec. per sec.	10.9	4.7	8.0	4.9	0.0001	27
5° per sec. per sec.	11.3	4.5	7.6	4.3	0.0001	33
12° per sec. per sec.	11.0	4.7	8.6	5.5	0.0002	22
Postrotation rate: nystagmi per second						
2° per sec. per sec.	1.16	0.46	0.81	0.46	0.0001	30
5° per sec. per sec.	1.27	0.47	0.97	0.52	0.0001	24
12° per sec. per sec.	1.27	0.47	1.03	0.56	0.0002	19

sity increased, the amount of deficit decreased in the more significant measures, acceleration latency and rotation rate. This inverse relationship between intensity and deficit is seen most clearly in acceleration latency. In the normal controls, the latency of acceleration nystagmus was nearly constant at the mild, moderate, and strong intensities used in the experiment.[4] The latencies of the schizophrenic group, on the other hand, became shorter and approached those of the controls.

Deceleration latency is a special instance of reactivity in which the schizophrenic group showed no significant deficit. Once responding to acceleration stimulation, both groups reversed their reaction to deceleration with equal facility. After reversal had taken place, however, the schizophrenic group at first (during rotation) showed least deficit to the moderate stimulus. In keeping with the general pattern, postrotation deficits were least to the strong stimulus.

Postrotation nystagmus is a phenomenon of receptor adaptation and recovery (Steinhausen, 1933). This after response corresponds to afterimages of rotation which are analogous to afterimages in other sensory modalities. The data in Table 2 amplify in terms of known stimulus intensities the finding of Angyal and Blackman (1941) of postrotation nystagmic deficit in schizophrenia.[5]

[4] Differential sensitivity of the right and left labyrinths may account for the slightly longer latency found for the controls at 5° per sec. per sec. than at 2° and 12° per sec. per sec. Rotation was clockwise at 5° per sec. per sec., which ipsilaterally stimulates the right semicircular canals. At 2° and 12° per sec. per sec., the left semicircular canals were stimulated ipsilaterally.

DISCUSSION

The direct relationship of nystagmus to the physiology of the semicircular canals has been shown by Steinhausen (1933), Dohlman (1935–36), and Lowenstein and Sand (1940). Their work established angular acceleration as the stimulus. In any head or body movement, eye fixation is maintained by semicircular canal reflexes. In the experimental situation, the fast phase has an adaptive function. Following the eye's slow deviation, the fast phase re-establishes the original fixation. This occurs when the visual field rotates in a constant relationship to the *S*.

An idea of the complexity of the process involved can be gained from a consideration of the events in experimental rotation. Acceleration occurs during the first part and deceleration during the last, when the direction of the slow and fast phases of nystagmus reverses. Deceleration nystagmus perseverates after stimulation (postrotation nystagmus). In a clockwise rotation, the appropriate sensation of rotation to the right is experienced during acceleration. In the brief interval when the stimulus changes from acceleration to deceleration, the rotation seems to stop. There follows the illusion of rotating in the opposite direction. This illusion continues into the postrotation period concomitant with postrotation nystagmus.

[5] See also Angyal (1941, 1940, 1942).

The results indicate schizophrenic deficits in each stage of this experimental sequence except deceleration latency. They do not indicate where in the neural chain of events (e.g., the semicircular canals) impairment, if any, might lie. The most conservative inferences suggested are those related to the complexity and integration of the neural processes.

1. Of foremost importance is the integrative or adaptive nature of the neural deficit. Considering the proprioceptor and orienting roles the semicircular canals play, the nature of the deficit suggests a biological concomitant of personality disturbances similar to those associated with the body image.

2. The results also emphasize the dimensional nature of deficit in schizophrenia. Nystagmic reactivity for both groups falls into fairly normally distributed measures, and their standard deviations reveal considerable overlap.

3. Adequate neural function was demonstrated in at least one aspect of schizophrenic adaptation. When the stimulus or stress reversed from acceleration (positive) to deceleration (negative), no significant deficit in latency of deceleration nystagmus seemed to occur. Various social and personal aspects of schizophrenia may be subject to a similar principle.

4. With increasing intensity of the stimulus, schizophrenic deficit decreased. This tendency is demonstrated most clearly in latency of nystagmus to onset of rotation. Schizophrenic deficit was less than one-fourth for moderate and strong intensities than for a mild intensity. In its more general implications, this relationship is suggested by stress concepts of Seyle (1956). The greater the intensity of the stress, the closer organisms may approach the uniformity of the general adaptation syndrome.

5. Confirmation of findings of Angyal and Blackman (1940, 1941) of reduced postrotation nystagmus in schizophrenia was experimentally established. Their postrotational mean rates of 1.32 for the patients and 1.68 for the normal controls show a schizophrenic deficit of 21%. Even though the deceleration stimulus produced by the Barany method is estimated at upwards of 1000° per sec. per sec., post rotation deficits are greater at 2° and 5° per sec. per sec. (30% and 24% respectively).

The procedure was especially applicable to mentally ill *S*s with communication difficulties. Minimal cooperation entirely of a nonverbal nature was required of the *S*s. The automatic nature of the response prevented *S*'s rapport from affecting the measurements.

Other advantages were provided by the experimental conditions. None of the *S*s were apparently bothered by aftereffects of rotation. The functional relationship between stimulus intensity and nystagmus could be established because both were precisely quantified. Noteworthy is the comparison of the Barany method to that employed in the experiment reported here. Stimulus levels far below the shock level of that clinical technique were more adequate in identifying the degree of neural integration.

SUMMARY

This experiment demonstrated some aspects of neural integrative adequacies and deficits in schizophrenic function. The nystagmic reflex is a useful measure because of its adaptive nature related to an individual's thresholds, visual-motor orientation, and proprioception. Certain conclusions and implications may be enumerated.

1. Schizophrenics as a group have deficits in latency and amount of nystagmus varying with intensity of stimulation.

2. Adequacy of schizophrenic integration and sensitivity increased as intensity of stimulation increased.

3. No deficit occurred when the stimulus reversed from positive to negative (acceleration to deceleration).

4. A mild stimulus was more effective in demonstrating deficit than stronger stimuli.

REFERENCES

ANGYAL, A. The perceptual basis of somatic delusions in a case of schizophrenia. *Arch. Neurol. Psychiat.*, 1935, **34**, 270–279.

ANGYAL, A. The experience of the body-self in schizophrenia. *Arch. Neurol. Psychiat., Chicago*, 1936, **35**, 1029–1053.

ANGYAL, A. Disturbances of activity in a case of schizophrenia. *Arch. Neurol. Psychiat., Chicago*, 1937, **38**, 1047–1054.

ANGYAL, A., & BLACKMAN N. Vestibular reactivity in schizophrenia. *Arch. Neurol. Psychiat., Chicago*, 1940, **44**, 611–620.

ANGYAL, A., & BLACKMAN, N. Paradoxical reactions in schizophrenia under the influence of alcohol,

hyperpnea, and CO_2 inhalation. *Amer. J. Psychiat.*, 1941, **97**, 893–903.

ANGYAL, A., FREEMAN, H., & HOSKINS, R. G. Physiologic aspects of schizophrenic withdrawal. *Arch. Neurol. Psychiat.*, 1940, **44**, 621–626.

ANGYAL, A., & SHERMAN, N. A. Postural reactions to vestibular stimulation in schizophrenic and normal subjects. *Amer. J. Psychiat.*, 1942, **98**, 857–862.

DOHLMAN, G. Toward a method for quantitative measurement of the functional capacity of the vestibular apparatus. *Acta oto-Laryng., Stockh.*, 1935–36, **23**, 50–62.

HOLSOPPLE, J. Q. Space and the non-auditory labyrinth. In C. Murchison (Ed.), *Foundations of experimental psychology*. Worcester: Clark Univer. Press, 1929, Pp. 414–433.

LEACH, W. W. The measurement of nystagmus in normals and schizophrenics. *Dissertation Abstr.*, 1953, **13**, 598–599. Publ. No. 5393, (Univer. Microfilms, Ann Arbor, Mich.)

LOWENSTEIN, O., & SAND, A. The individual and integrated activity of the semicircular canals of the elasmobranch labyrinth. *J. Physiol.*, 1940, **99**, 89–101.

MOWRER, O. H. Concerning the normal function of the vestibular apparatus. *Ann. Otol., Rhinol., & Laryngol.*, 1932, **41**, 412–421.

MOWRER, O. H. The nystagmic response of the pigeon to constant angular acceleration of liminal and supraliminal intensities. *J. comp. Psychol.*, 1935, **19**, 177–193.

SELYE, H. *The stress of life*. New York: McGraw-Hill, 1956.

STEINHAUSEN, W. Über die Beobachtung der Cupula in den Bogengangsampullen des Labyrinths des lebenden Hechts. *Pflueg. arch. ges. Physiol.*, 1933, **232**, 500–512.

TER BRAAK, J. W. G. Über die Empfindlichkeit des Bogengangsapparates für Winkelbeschleunigungen. *Pflueg. arch. ges. Physiol.*, 1936, **238**, 319–326.

(Received November 7, 1958)

Journal of Abnormal and Social Psychology
Vol. 60, No. 3, 1960

SELF-ESTEEM AND NEED ACHIEVEMENT AS DETERMINANTS OF SELECTIVE RECALL AND REPETITION[1]

STANLEY COOPERSMITH
Wesleyan University

PERSONALITY theorists ascribe the selectivity of memory to the screening of experiences in terms of their acceptability to the ego. Stimuli that threaten the person's self-esteem are assumed to be less available to consciousness while ones that are acceptable or enhancing are permitted freer expression. Anxiety, threat, stress, and involvement are terms that have been used to describe the psychological state which ensues when there is the possibility of lowered self-esteem. Psychological defenses against anxiety maintain as positive a self-evaluation as is possible, and, as Freud notes, "defenses proceed from the self-respect of the ego" (Freud 1924). Studies of success and failure have suggested (Alper, 1957; Eriksen, 1952, 1954) that where self-esteem is firmly established, failures are more readily recalled and perceived, while insecure or lowered self-esteem results in denial and heightened thresholds for negative stimuli. Similar investigations relating to achievement motivation (Alper, 1957; Atkinson, 1953) strongly suggest that the desire to overcome a failure is a further influence upon the direction of recall, and one that is likely to be expressed in behaviors designed to overcome the affect thereby aroused. The purposes of the present experiment were to investigate the relationship between several indices of self-esteem, achievement motivation, and the recall and repetition of success and failure experiences; and examine the total and positive and negative aspects of n Achievement within the context of self-esteem types previously found to be reliable and consistent (Coopersmith, 1959), and determine the manner in which these aspects are related to these experimental behaviors.

Task interruption has been widely used in experiments concerning selective recall. Zeigarnik (1927) introduced the method with the finding that incompleted tasks were recalled more frequently, and subsequent studies have attempted to determine the variables associated with different recall ratios. Converging lines of evidence now indicate that the recall pattern is a function of the stress introduced into the situation, individual differences in ability to tolerate the stress, and motivations aroused by its introduction. Studies of group trends (Alper, 1946; Glixman, 1949; Rosenzweig & Mason, 1934; Rosenzweig, 1943), although inconsistent, tend to indicate that as threat to self-esteem increases, the recall of incompleted tasks decreases, while the recall of completed tasks remains fairly constant. A number of early investigators remarked upon the personality characteristics of subjects who recalled more failures. They were variously described as proud (Rosenzweig & Mason, 1934), ambitious (Zeigarnik, 1927), and needing success (Cartwright, 1942), and older (Sanford, 1946), and more mature (Rosenzweig & Mason, 1934). Recent studies, more empirical and dynamic in formulation, have investigated personality dimensions broader and more basic in nature. Recall of failure under stressful conditions has been reported for persons high in ego strength (Jourard, 1954; Tamkin, 1957) and achievement motivation (Atkinson, 1953), who show relatively little perceptual defense (Eriksen, 1952) and are more sure of their personal and intellectual worth (Caron & Wallach, 1957; McArthur, 1954). The interaction of situational and personality variables results in four possible recall patterns, two of which, (*a*) a shift from failures to successes under stress, and (*b*) a shift from successes to failures under stress, have been reported. The two other patterns (*c*) recall of successes under all conditions, and (*d*) recall of failures under all conditions, have been suggested (Alper, 1948) but not reported.

In the present experiment subjects differing in subjective and behavioral evaluations of

[1] The research cited in this paper originated under USPHS Grant M (901) and is currently supported by USPHS Grant M(2696). An earlier version of this paper was presented to the 1959 meeting of the American Psychological Association.

self-esteem, and representing different types of self-esteem (Coopersmith, 1959), were individually presented a series of tasks, half of which were failed. A measure of achievement motivation was then obtained under what were designed to be stressful and arousing conditions. Selective recall and the subject's decisions to repeat either successes or failures were then elicited.

Repetition behavior was obtained on the assumption that it is a more concrete confrontation with previous experience than is recall, and may thus be more anxiety provoking. Furthermore, repetition involves a commitment of future behavior and thus provides an opportunity for testing the hypotheses (Caron & Wallach, 1957; Osgood, 1953) that recall is a cognitive precursor to behaviors that the person has previously found to be constructive. Repetition differs from resumption (Ovsiankina, 1928) in that it controls for differences in point of interruption on failed tasks, and provides an opportunity to redo all tasks in their entirety without distinction between successes and failures.

METHOD

Subjects

The sample consisted of 48 middle middle- and upper middle-class children enrolled in the fifth and sixth grades of an urban elementary school.[1] Four groups, of 12 *Ss* each, were chosen from a group of 102 children on the basis of their responses on the Self-Esteem Inventory (SEI), and a teacher's rating of their self-esteem behavior as recorded on the Behavior Rating Form (BRF). The four groups were found to differ significantly on a constellation of variables (achievement, sociometric status, ideal self, and n Achievement) and may be briefly characterized as follows: *High-High* (SEI-BRF)—above average performance commensurate with goals; *High-Low*—low achievers with high goals; *Low-Low*—low achievers who have accepted their inferior performance and position; and *Low-High*—high achievers with extremely high goals but with low self-esteem. It thus becomes possible to analyze patterns of recall and repetition not only in terms of the four types of self-esteem, but also to compare the results of groups high (H-H, H-L) and low (L-L, L-H) in subjective evaluation, and groups rated high (H-H, L-H) and low (H-L, L-L) on their self-esteem behaviors. The measures and methods used to select these *Ss* and the elaboration of the self-esteem types have been presented in a previous paper (Coopersmith, 1959).

In deciding which tasks should be employed, the following criteria were established: (*a*) equal attractiveness; (*b*) capable of accomplishment by every child; (*c*) requiring approximately the same amount of time; (*d*) generally familiar; (*e*) interesting; (*f*) success or failure related to the *S*'s performance; and (*g*) relative unlikelihood that *Ss* would interpret their accomplishment in terms of partial success and failure. In general, it was found that tasks requiring simple motor coordination were most effective.

In order to control for recall or reptition due to the nature or attractiveness of the task, a success and failure variation of each task was employed. Only those tasks rated approximately equal in attractiveness were used. Although this use of alternates of the same task would tend to increase the likelihood that the first would act as a mnemonic to the second, the tasks were sufficiently different and presented in such an order and manner that their relatedness was not clearly apparent. The six pairs of tasks consisted of: (*a*) Mosaics—*S* was asked to replicate a design of a boy and a flowerpot using plastic heptagonal blocks; (*b*) Cut-outs—*S* was shown a paper cut-out of an animal after which the *E* disassembled it and asked the *S* to reproduce it; (*c*) Scrambled sentences—*S* was asked to arrange the words presented to him on cards into a meaningful sentence; (*d*) Kohs blocks—*S* was asked to reproduce a design of a house and a series of stripes; (*e*) Jig-saw puzzle—*S* was shown a completed puzzle after which the *E* disassembled it and asked the *S* to reassemble it; and (*f*) Snap-block construction—*S* was asked to reproduce a design of a horse and a robot using varied shaped blocks that snapped together.

Procedure

Task administration. The set of 12 tasks was administered individually to each *S* under conditions designed to threaten self-esteem. *E* had previously announced to the class that they were to participate in a study designed to test the attitudes and achievement of children. The selected *S* was taken to a private office in the school building, seated before a desk, at the side of which sat *E*, stopwatch in hand. *S* was informed that he would be permitted two minutes to complete each task and that *E* would time him and record the nature of his performance. The tasks were out of *S*'s sight until *E* presented them individually in the sequence S F S S F F S F F S S F. They were rotated so that success tasks occurred an equal number of times in each success position and failure tasks an equal number of times in each failure position. Success and failure were controlled by timing, with the difference between success and failure variations of each task generally less than 30 sec. Analysis indicated no effect due to either serial position or point of interruption. At the conclusion of each task *E* announced that *S* had either succeeded or failed upon the task; he repeated this before proceeding with the next task.

Measurement of n *Achievement.* After *E* had placed the last task out of view, he asked *S* to complete four stories that had previously been found (McClelland,

[1] The author wishes to thank the superintendent, principals, faculty, and students of the Ithaca Public School System for their cooperation with this study.

Atkinson, Clark, & Lowell, 1952) to elicit n Achievement in children. This measure was included to determine motivation under arousal conditions. It provided a "filler" between performance and recall, and made it possible to evaluate the relationship between achievement motivation and recall as previously demonstrated (Atkinson, 1953).

The story completions dealt with: a boy (or girl) at school; brothers and sisters playing; two men (or women) near a machine; and a boy (or girl) seated at a desk. As the child presented the stories orally, they were recorded by *E*. The average duration of this period was approximately 15 min. It had previously been determined (McClelland et al., 1952) that stories were more successful than pictures in eliciting achievement motivation from children.

The scoring of the stories followed the established criteria (McClelland et al., 1952), each story being scored separately for the presence or absence of each of the 13 achievement related categories. They were further grouped into the *positive* (I+, Ga+, G+, Achievement Thema) and *negative* aspects (I−, Ga−, G−, B) subsumed under the total n Achievement score, aspects which are apparently indicative (McClelland et al., 1952) of hope of success and fear of failure respectively. A product-moment correlation of .85 was obtained between the total n Achievement scores of two independent judges.

Obtaining recall and repetition. After *S* had completed the last of his stories, *E* asked him to report which of the tasks he remembered most clearly of all upon which he had previously performed, then as many others as he could recall. Pretests had established that a large percentage of the children could recall all the tasks and it was therefore decided to include a measure of 'clearest" recall, with instructions indicating the acceptability of any recall preference.

E then led the child behind a curtain to where the tasks, all completely assembled, were located. He indicated a pair of tasks, offered *S* an opportunity to repeat either the success or failure variation, and after the decision had been made allowed him to redo the task. Decisions were obtained for the other pairs of tasks, although *S* did not actually undertake the task again in the remaining five cases. Before leaving he was asked to keep the entire procedure in confidence. Insofar as is known, no breach of confidence occurred.

TABLE 1

MEAN NUMBER OF SUCCESSES RECALLED BY DIFFERENT TYPES OF SELF-ESTEEM

Groups[a]	Initial	First Three[b]	Total[b]	Fail/ Succ. Ratio
Low-Low	0.8	2.2	4.7	.84
Low-High	0.5	2.2	5.1	.89
High-Low	0.5	2.0	4.8	.91
High-High	0.8	2.3	5.3	.83
P[c]	<.30	>.30	>.30	—

[a] The terms refer to upper vs. lower quartile status on the Self-Esteem Inventory and the Self-Esteem Behavior Rating Form, respectively.

[b] The individual scores for the first three, and total recalls were transformed into percentage success form. The recall percentages were then transformed into angles by the arc sine percentage transformation.

[c] *P* is based on χ^2 distribution for initial recall, and on the F distribution for the first three and total recalls.

TABLE 2

FREQUENCY OF RECALL AND REPETITION PATTERNS FOR SELF-ESTEEM TYPES

Group	Recall			Repetition		
	More Fail.	Equal	More Succ.	More Fail.	Equal	More Succ.
Low-Low	2	4	6	1	2	9
Low-High	1	5	6	9	2	1
High-Low	3	3	6	8	2	2
High-High	0	5	7	1	4	7

RESULTS

Self-Esteem

In Table 1 the recall behaviors of the different types of self-esteem are compared. In order to test for a possible sequential effect, initial recall, first three recalls, and total recall of successes were noted. The most striking feature of these results is the lack of any significant differences between the self-esteem groups. The Low-High and High-Low groups showed a tendency to recall more failure on the first recall but this difference is well within chance expectancy. On the first three recalls all groups recalled at least twice as many successes as failures, while the ratio of failure to success of less than one for total recall corroborates the general tendency for the recall of success tasks. The possibility that *S*s in each group would differ in the frequency with which they recalled more successes, more failures, or an equal number of each is refuted by Table 2. Chi square for the distribution of *S*s falling into each category was clearly not significant.

Comparison of the groups high (H-L, H-H) and low (L-H, L-L) in subjective evaluation reveals no differences on any of the recall measures, nor does comparison of the groups high (H-H, L-H) and low (L-L, H-L) on ratings of self-esteem behavior. The present findings, then, do not support the view that selective recall is related to subjective or behavioral indices of self-esteem, nor do children representing different types of self-esteem differ in their recall patterns. These results are consistent with previous studies

TABLE 3
CORRELATIONS OF MEASURES RELATED TO RECALL AND REPETITION

Variable	No. of Successes Recalled	No. of Successes Repeated
Number of successes recalled	—	−.26
Self-Evaluation (SEI)	−.21	−.25
Behavior Evaluation (BRF)	−.03	.13
Objective Achievement (Iowa)	−.08	.23
D(SB)[a]	−.08	−.27
Manifest Anxiety	.08	−.11
Ideal Self	.02	−.17
Need Achievement	.12	−.38
D(SI) adjusted[b]	.16	.07

Note.—Value of r for $N = 48$, .05 level = .28, .01 level = .37.
[a] Difference between Self and Behavioral Evaluations.
[b] Difference between Self and Ideal Self, adjusted for "ceiling" effect.

(Caron & Wallach, 1957; Glixman, 1949; Smock, 1957) in the finding that there is a greater recall of successes than failures under stressful conditions.

For task repetition, the dependent variable is the total number of success tasks repeated. Table 2 indicates that *Ss* in the Low-High and High-Low groups tended to repeat failures while Low-Lows and High-Highs repeated successes. Thus, groups showing agreement between SEI and BRF scores repeated twice as many successes as failures, while cases of disagreement repeated failures twice as often as successes. The comparison of groups high and low in self-evaluation, and the groups high and low on their behavioral ratings show no differences in repetition.

In the course of preparing the profiles of the self-esteem types (Coopersmith, 1959) information on several personality and motivational variables had been obtained. A correlational analysis was performed to determine the nature and extent of the relationships between these measures and the recall and repetition of successes. The results of this analysis are presented in Table 3. The analysis includes data based on three scores derived from differences between measures. The first of these is the D(SB) score, based on the difference between subjective and behavioral evaluations. Since the groups with high positive (H-L) and high negative (L-H) D(SB) scores behaved in identical fashion in the repetition experiment, the direction of the discrepancy will not be considered in this analysis. The D(SI) and the D(SI) adjusted scores are based upon the difference between subjective and ideal self-statements, with the latter score adjusted to correct for the ceiling imposed on the D(SI) score by persons high in their self evaluation. Since the correlation between the two D(SI) scores was significant ($r = .64$, $p < .01$) only the adjusted score is included.

None of the variables in Table 3 appear to be significantly related to the number of successes recalled. The only correlation approaching significance was between the number of successes recalled and number of successes repeated ($r = -.26$, $p < .08$). Of the four other variables related to the repetition of successes, achievement motivation was clearly significant ($p < .01$), while the D(SB) score ($p < .06$), self evaluation ($p < .09$), and objective achievement ($p < .10$) approached accepted significance levels. Partial correlations indicate that n Achievement and the D(SB) score, the two best predictors, are each independently related to repetition behavior. The beta coefficients of the multiple correlation of repetition with the n Achievement, D(SB), and SEI scores combined ($r = .49$, $p < .01$), signify that these variables contribute respectively 45%, 34%, and 21% of the relationship. The coefficient of multiple determination ($R^2 = .2349$) indicates that these variables combined can account for 23% of the variation in repetition.

Achievement Motivation

The finding that the two self-esteem types (High-Low and Low-High) that are high in achievement motivation but significantly different in other characteristics, both repeat failures, raises the question of whether their actions are based upon similar underlying motives or upon different motives which produce the same effect. A number of studies (Alper, 1957; Atkinson, 1953; McClelland et al., 1952) indicate that the total n Achievement score includes a positive aspect indicative of an approach orientation and a motivation to overcome obstacles, and a negative aspect oriented toward avoiding failure. On the basis of the curvilinear relationships found in studies of perception (McClelland et al., 1952) and recall (Atkinson, 1953), it has been proposed (McClelland et al., 1952) that negative re-

sponses occur with greater relative frequency in the middle and lower portions of the distribution of total n Achievement scores.

Examination of Table 4 indicates that there are significant differences between the groups in positive and negative subcategories as well as in the general categories. While all groups have more responses in the positive than in the negative categories, the Negative/Positive ratio indicates that the High-Highs have the highest proportion of positive responses (and are thus presumably approach-oriented), the Low-Lows the highest proportion of negative responses (and are thus presumably avoidant-oriented), while the two groups showing discrepancies between self-ratings and behavior have more of a balance between positive and negative categories, a situation which would presumably lead to conflict. These findings are consistent with the personality characteristics and dynamics of the self-esteem types (Coopersmith, 1959) and support the belief that they represent distinct character structures.

Inspection of Table 4 further reveals that of the two groups which are low in achievement motivation the Low-Low self-esteem group shows the greatest percentage of negative responses, whereas the High-Highs show the greatest percentage of positive responses. The findings indicate that the positive and negative categories may vary independently of each other and of the total n Achievement score.

TABLE 4

TOTAL NUMBER OF RESPONSES OF THE SELF-ESTEEM GROUPS IN GENERAL CATEGORIES AND POSITIVE AND NEGATIVE SUBSCORES

Group	General Categories			Total n Ach and Subscores			Ratio, Neg./ Pos.
	AI[a]	TI[b]	UI[c]	Total	Negative	Positive	
Low-Low	22	15	11	48	16	19	.84
Low-High	33	5	10	105	26	47	.55
High-Low	23	12	13	72	21	34	.62
High-High	19	12	17	49	8	32	.25
	*	*	—	**	*	*	**

* $p < .05$.

** $p < .01$.

[a] Achievement Imagery—Indicated by long term involvement, unique accomplishment, or competition with a standard of excellence.

[b] Task Imagery—Some reference to achievement but not explicit in terms of AI criteria.

[c] Unrelated Imagery—No references to achievement.

TABLE 5

POSITIVE AND NEGATIVE SUBCATEGORIES OF ACHIEVEMENT MOTIVATION BY SELF-ESTEEM GROUPS

Positive	I+	Ga+	G+	Ach Th	Tot. Pos.	Pos–Ach Th
Low-Low	9	3	7	0	19	19
Low-High	14	8	14	11	47	36
High-Low	12	6	11	5	34	29
High-High	10	3	11	8	32	24
	—	—	—	*	*	—

Negative	I–	Ga–	G–	Bw	Bp	Tot. B	Tot. Ng	Neg.–Bw
Low-Low	0	3	11	1	1	2	16	14
Low-High	—	6	19	0	1	1	26	26
High-Low	2	4	7	6	2	8	21	15
High-High	0	1	6	1	0	1	8	7
	—	—	—	*	—	*	*	*

* $p < .05$.

** $p < .01$.

Note.—Key to subcategories:

Positive

I+ Successful instrumental activity

Ga+ Anticipation of success

G+ Positive affect related to achievement

Ach Th Achievement imagery central theme

Negative

I– Unsuccessful instrumental activity

Ga– Anticipation of failure

G– Negative affect related to achievement

Bw Environmental obstacle

Bp Personal obstacle

With respect to responses within the subcategories, Table 5 indicates that the major difference for the positive subcategory is in Achievement Thema responses while that of the negative aspect is in Bw responses representing impersonal obstacles to achievement. The difference in the negative category remains significant when the Bw responses are removed, whereas the difference in the positive category no longer holds with the removal of Achievement Thema responses.

In order to take account of the differences between the groups in Achievement related stories, the total positive and negative categories as well as the specific responses for each *S* were divided by the number of his AI stories. This analysis indicated significant differences for only the total negative category and the Achievement Thema and Bw responses.

Correlational analysis of the achievement motivation subcategories with recall and repe-

tition revealed a significant negative relationship between total positive responses and the number of successes repeated ($r = -.36$, $p < .01$). The correlation for the total positive alone ($r = -.36$) is similar to that attained for the total n Achievement score ($r = -.38$). None of the other correlations between categories and responses and the experimental behaviors, whether by type or for the total group, achieved accepted levels of significance, thus raising the possibility that where differences were obtained they could be accounted for by chance. A further limit on interpreting the relationship between n Achievement and repetition is set by the recognition that the groups which were high in their D(SB) scores were also high in n Achievement, thus raising the possibility that the findings could be due to the effects of either variable, of neither, or of yet another variable as yet unidentified. Interpretation must therefore await experimental evidence indicating the significance of the self-behavioral discrepancy, and the independent effects of the n Achievement and D(SB) variables.

DISCUSSION

The results of the present study clearly do not support the hypothesis that either achievement motivation or self-esteem are related to recall preference. The general result indicating a greater recall of successes under stress is in agreement with the results of previous studies (Glixman, 1949; Rosenzweig, 1943; Smock, 1957), despite a bias introduced in the sample selection. Of those cases where a greater recall of failures under stress has been reported, only one (Atkinson, 1953) indicates an absolute majority of failures, while the others refer to either the recall pattern of particular individuals (Alper, 1946; Atkinson, 1953; Caron & Wallach, 1957; Eriksen, 1952) or to an increase relative to some previous finding (Alper, 1948; Glixman, 1949; Rosenzweig, 1943; Smock, 1957). Thus, while the suggested relationships of this study are not borne out, the group results support the general finding of a recall bias favoring success.

It is this very absence of relationships between the battery of variables and recall, as well as the small proportion of individuals in the sample who recall more failures, which raises serious yet intriguing questions. Turning first to situational and task variables for an explanation, it may be argued that *S*s in the present experiment may have been placed under greater stress than is generally the case, and since one of the effects of stress is to decrease the number of failures recalled (Glixman 1949; Rosenzweig, 1943; Smock, 1957) superordinate stress should presumably accentuate this trend. While this may have been the case, *E*'s observations and the comments by the *S*s did not indicate an arousal greater than that reported by others investigators. In addition, recall experiments differ on such variables as the number of tasks used (4–30), the method of administration (group or individual), the type of task (scrambled sentences, drawing, puzzles, problems, etc.), the length of tasks, the length of time between performance and recall, and whether or not a "filler" activity is introduced.

A further reason that may be advanced is that the use of alternates of the same task may have acted as a mnemonic and thereby confounded the recall pattern by increasing it *in toto*. Here the matter would seem to be one of strategy since it is clearly necessary to control for the nature and attractiveness of the task, and while each of the tasks may well act as a mnemonic for the other, procedures which do not control this variable are subject to the contrary rejoinder. While these differences make comparisons difficult, it should be noted that the procedures employed in the present study are essentially similar to those employed in studies reporting a much larger percentage of *S*s recalling more failures than successes (Alper, 1946; 1957; Caron & Wallach, 1957; Glixman, 1949; Jourard, 1954; Tamkin, 1957).

The most likely explanation for the present findings appears to be that the *S*s were younger than those generally employed in recall studies. It is clearly not the passage of time but rather the quantity of experiences as well as their quality (success and failure) which is presumably related to recall preference. The two characteristics linked to maturity that appear to be related to recall are: (*a*) the development of a tolerance for the anxiety aroused by failure and a perspective upon its significance, an ability subsumed under such terms as ego

strength and self-esteem and (*b*) an attitude towards failure as frustrating and insulting which results in behavior designed to change it to success, a measure of which is apparently obtained in the achievement motive. Persons high in both these characteristics should recall failures under stress and nonstress, while those low in both these characteristics should recall successes under both conditions. Persons with intermediate amounts of the two, in varying combinations, may be expected to recall either success or failure, depending on the extent to which the situation is perceived as failure and the extent to which n Achievement is aroused by situational cues.

A good deal of the controversy as to whether the recall of failures is indicative of a weak (Alper, 1946; Rosenzweig & Mason, 1934) or strong ego (Sanford, 1946) appears due to a failure to distinguish between the concepts of *appropriateness* and *strength*. Appropriateness refers to a response suitable to a particular situation and may be acquired through relatively superficial social learning, while psychological strength refers to the ability to tolerate adversity and less need to distort and restrict reality, thus increasing the probability that adversity will be overcome. The former is based upon familiarity with a given environment and perceptual sensitivity to its cues, while the latter is achieved by states of high motivation, a wide range of experiences, and the confidence gained in overcoming adversity. The two clearly are not synonymous and it is plainly not possible to reason from one to the other.

Failure to distinguish between the appropriateness and strength in the case of recall leads to the position that the distortion and restriction of reality which results from protecting self-esteem is an indication of greater strength than accurate perception and a corrective response. It also leads to the conclusion that children have stronger egos than do adults, a conclusion clearly unwarranted by current theory or evidence. A more meaningful interpretation would appear to be that adults have learned to act more appropriately than children. Accordingly, they may be expected to report responses publicly in accord with group expectations, ones that do not necessarily correspond to the total or most vivid recall experienced.

In evaluating the hypothesis (Caron & Wallach, 1957; Osgood, 1953) that recall is a rehearsal of future behavior with a view towards commitment, it is necessary to distinguish between those cases of repetition in which the *S* is confronted with the various tasks and asked to choose between them, and those in which the repetition choices must come from the tasks which he has recalled. The former case allows a wider psychological world to operate than does the latter, which may provide alternates for repeating none of the variables, only one, or a choice among several variations. It is obvious that if the person is motivated to overcome a failure experience but cannot recall it, repetition is an impossibility unless he is confronted by it. Since the procedures for aided and unaided repetition may produce different results and the present test of the rehearsal hypothesis is limited to the condition of aided repetition, it may be suggested, in view of the relatively low correlation ($r = .26$) between recall and repetition, that if recall is a rehearsal, it is not one that frequently results in performance. Since the results indicate that more successes are recalled than failures, unaided repetition, i.e., that based solely on the recall of the *S*, would appear to be more likely to result in a greater repetition of successes than would aided repetition.

It should be borne in mind that aided repetition is more anxiety provoking since it forces upon the individual the recognition that he has failed, something which can be denied when the defenses are allowed full sway. In terms of failure repetition, which would appear under all circumstances to cope more effectively with reality, the Low-High and the High-Low groups apparently are more capable and willing to undergo anxiety in order to attain success than are the High-High and Low-Low *S*s who continue to repeat successes even when confronted with failures.

SUMMARY

This study investigates the relationship between self-esteem and the recall and repetition of success and failure experiences. Four groups, of 12 *S*s each, were selected on the basis of their self-evaluative responses and an evaluation of their self-esteem behaviors. The combinations of these two variables, at their

extremes, yielded significantly different patterns on such variables as achievement, ideal self, and sociometric status, and apparently represent distinct types of self-esteem.

Each *S* was given a series of six pairs of tasks, in an achievement oriented situation. The *S* failed one of each pair of tasks and succeeded on the other. The tasks were followed by four story completions which had been found to elicit n Achievement. Recall was then obtained, after which the *S* was offered an opportunity to repeat either the success or failure variation of each task.

The data were analyzed both in terms of the four groups differing in self-esteem and by a correlational analysis to determine which variables were related to recall and repetition. All groups recalled more successes, and none of the independent variables was significantly related to recall. Groups in which self-rated self-esteem was discrepant from behavior ratings tended to repeat failures; those in which the two measures agreed tended to repeat successes. An index of agreement between self-rated self-esteem and behavioral self-esteem and n Achievement were the variables most significantly related to repetition. Analysis of the need achievement scores indicated differences between the self-esteem types in positive, negative, and general categories as well as subcategory responses, although none of these were related to the experimental behaviors.

It is proposed that two factors are necessary for the recall and repetition of failure. These are the ability to tolerate failure and the motivation and striving to overcome its effects.

REFERENCES

ALPER, THELMA G. Memory for completed and incompleted tasks as a function of personality: An analysis of group data. *J. abnorm. soc. Psychol.*, 1946, **41**, 403–420.

ALPER, THELMA G. Task orientation and ego orientation as factors in reminiscence, *J. exp. Psychol.*, 1948, **38**, 224–238.

ALPER, THELMA G. Predicting the direction of selective recall: Its relation to ego strength and n Achievement. *J. abnorm. soc. Psychol.*, 1957, **55**, 149–165.

ATKINSON, J. Achievement motivation and the recall of interrupted and completed tasks. *J. exp. Psychol.*, 1953, **46**, 381–390.

CARON, A. J., & WALLACH, M. A. Recall of interrupted tasks under stress: A phenomenon of memory or of learning? *J. abnorm. soc. Psychol.*, 1957, **55**, 372–381.

CARTWRIGHT, D. The effect of interruption, completion, and failure upon the attractiveness of activities. *J. exp. Psychol.*, 1942, **31**, 1–16.

COOPERSMITH, S. A method for determining types of self-esteem. *J. abnorm. soc. Psychol.*, 1959, **59**, 87–95.

ERIKSEN, C. W. Defense against ego-threat in memory and perception. *J. abnorm. soc. Physchol.*, 1952, **47**, 230–235.

ERIKSEN, C. W. Psychological defenses and "ego strength" in the recall of completed and incompleted tasks. *J. abnorm. soc. Psychol.*, 1954, **49**, 45–50.

FREUD, S. On narcissism: An introduction. In *Collected papers*. Vol. 4. London: Hogarth, 1924.

GLIXMAN, A. F. Recall of completed and incompleted activities under varying degrees of stress. *J. exp. Psychol.*, 1949, **39**, 281–295.

JOURARD, S. M. Ego strength and the recall of tasks. *J. abnorm. soc. Psychol.*, 1954, **49**, 51–58.

MCARTHUR, C. Personalities of public and private school boys. *Harvard educ. Rev.*, 1954, **24**, 256–262.

MCCLELLAND, D. C., ATKINSON, J. W., CLARK, R. A., & LOWELL, E. L. *The achievement motive*. New York: Appleton-Century-Crofts, 1952.

OSGOOD, C. *Method and theory in experimental psychology*. New York: Oxford Univer. Press, 1953.

OVSIANKINA, MARIA. Investigations on the psychology of action and affection: VI. The resumption of interrupted acts. *Psychol. Forsch.*, 1928, **11**, 302–379.

ROSENZWEIG, S. & MASON, G. An experimental study of memory in relation to the theory of repression. *Brit. J. Psychol.*, 1934, **24**, 247–265.

ROSENZWEIG, S. An experimental study of "repression" with special reference to need-persistive and ego-defensive reactions to frustrations. *J. exp. Psychol.*, 1943, **32**, 64–74.

SANFORD, R. N. Age as a factor in the recall of interrupted tasks. *Psychol. Rev.*, 1946, **53**, 234–240.

SHOCK, C. D. Recall of interrupted and non-interrupted tasks as a function of experimentally induced anxiety and motivational relevance of the task stimuli. *J. Pers.*, 1957, **25**, 589–599.

TAMKIN, A. S. Selective recall in schizophrenia and its relation to ego strength. *J. abnorm. soc. Psychol.*, 1957, **55**, 345–349.

ZEIGARNIK, B. On finished and unfinished tasks. In W. D. Ellis (Ed.), *A source book of Gestalt psychology*. New York: Harcourt Brace, 1938, Pp. 300–314.

Journal of Abnormal and Social Psychology
Vol. 60, No. 3, 1960

COOPERATIVE BEHAVIOR IN DYADS AS A FUNCTION OF REINFORCEMENT PARAMETERS[1]

SEYMOUR ROSENBERG

Bell Telephone Laboratories, Murray Hill, New Jersey

STUDIES of cooperation in the laboratory have been stimulated, in part, by the large number of life situations in which cooperative action is observed both to occur and to break down. Biologists and ethologists (Lindzey, 1954, pp. 540–541; Tinbergen, 1953) have isolated and analyzed a number of examples of cooperative behavior among the lower organisms. These relationships, termed symbiotic or commensal, have frequently been noted in dyads for feeding or protective purposes and are often composed of members from different species. Elaborate systems of cooperative behavior as well as breakdowns in these systems also pervade human activity (Lindzey, 1954, p. 754; Mintz, 1951). Many cooperative situations are difficult to analyze with currently available techniques and theory. The abstraction and isolation of variables in the laboratory should enhance our understanding of these complex phenomena. The purpose of this study is to describe a number of reinforcement parameters in terms appropriate to dyadic cooperation and to study these parameters experimentally. The choice of reinforcement parameters for the present study is based upon the general conception that the reinforcing events are critical components in the development and maintenance of cooperative action.

Even a cursory review of the literature on cooperation reveals a number of different usages of the concept (Deutsch, 1949). To avoid ambiguity, a definition of a cooperative situation is adopted for this study that does not attempt to include all possible uses of the term, but does share an area of meaning with other definitions in the literature and is intended to be applicable to a variety of experimental conditions. The definition is formulated for dyads but extensions to larger groups are easily made. A cooperative situation has all of the following four characteristics: (*a*) There exist at least two outcomes for each individual following a response. (*b*) At least one of these outcomes is rewarding (i.e., it provides positive reinforcement and/or terminates noxious stimulation), and at least one outcome is not rewarding (or less rewarding). (*c*) Reward delivered to one person must not preclude its availability for the other. In this study a reward was available to both persons simultaneously during each trial. A cooperative situation may be contrived, however, in which reward is deliverable to only one or the other person in any one trial or time period. If the study consists of only one trial, the latter situation would be termed competitive rather than cooperative. (*d*) The outcomes to both persons are, in part or entirely, a function of the response of the other person. That is, the outcome to Person X must depend to some extent upon the behavior of Person Y *and* the outcome to Person Y must depend to some extent upon the behavior of Person X. If the outcome to only one is dependent on the behavior of the other, it is only a "partially cooperative situation." It resembles a commensal rather than a symbiotic relationship. There may, however, be additional asocial determinants for the outcomes in a cooperative situation, e.g., a random component.

Cognitive concepts are not included in this definition of a cooperative situation. "Awareness," defined either in terms of experimenter operations (e.g., instructions to the *S* that it is a "cooperative situation") or in terms of the *S*'s behavior, (e.g., questionnaire to the *S* to determine his verbalizations about the cooperative aspects of the situation) is not a necessary condition for the specification of a cooperative situation. Nor is it always a necessary condition for the development and maintenance of cooperative action (Hall, 1957;

[1] This research was performed by the author at Lackland Air Force Base, Texas. This research was done under ARDC Project No. 7737, Task No. 27006, in support of the research and development program of the Air Force Personnel and Training Research Center, Lackland Air Force Base, Texas. Permission is granted for reproduction, translation, publication, use, and disposal in whole and in part by or for the United States Gov.

Rosenberg & Hall, 1958; Sidowski, Wyckoff, & Tabory, 1956; Sidowski, 1957). That is, the consistent occurrence of response combinations that produce optimally rewarding outcomes for both persons does not require awareness in either sense of the term. The exclusion of awareness as a necessary condition for the definition of a cooperative situation, however, implies nothing about its relevance to the behaviors that take place in some cooperative situations.

The experimental situation adopted in this study has been used in a number of previous studies (Hall, 1957; Rosenberg, 1959; Rosenberg & Hall, 1958). While these studies were concerned primarily with the development and maintenance of precise motor activity in a team setting, the task and procedures were also found to be quite applicable to the problems discussed in this paper. Each person was given a series of trials in which he performed the simple motor response of turning a small concealed knob. It is a trial-and-error response in that the only discriminative stimulus for each trial was a signal to respond. The reinforcing stimulus was delivered to each *S* after each trial. Selected parameters of these reinforcing stimuli were considered important to the development and stability of cooperative action. These parameters were varied experimentally in a series of three experiments. The first experiment deals exclusively with cooperative situations. The other two experiments contain some conditions that are only partially cooperative. All conditions and every trial, however, involved a combination of outcomes that was positively reinforcing to both *S*s. The degree to which *S*s were able to learn to procure these simultaneous reinforcements, i.e., develop cooperative action, was studied.

The task used in these studies restricts and simplifies the interpersonal actions that occur to a much greater degree than has been customary in social psychology. Special emphasis has been placed on controlling the way in which one person's responses may affect another's stimulation. The control of these linkages appears to be central to extensions of S-R descriptions to interpersonal events, at least when viewed from our present state of understanding. Similar attempts to extend S-R theory have been expressed in the experimental studies of Sidowski et al. (1956) and Sidowski (1957) in their use of a "minimal social situation" and in earlier discussions by Miller and Dollard (1941), and by Logan (1955).

METHOD

Apparatus

Two identical adjacent booths were provided, one for each *S*. The walls of the booth were sufficiently high so that *S*s could not see each other. One wall of each booth consisted of a panel from which a small cylindrical knob projected about ½ in. Each *S* was provided with his own panel and knob. The knob was mounted about 3 ft. above floor level and was concealed by a wooden shield. The *S* could turn the knob by reaching under the shield but had no visual access to the knob. The knob was homogeneously knurled on its entire surface to avoid giving tactual cues of the amount it was turned. The *E*, seated on the other side of the panel, could read from a dial the amount each *S* turned the knob on each trial.

On the panel in front of each *S*, about a foot above the shielded knob, two or three 6-w. lights (depending upon the experimental condition) were mounted about 1.5 in. apart and labeled so as to provide a reinforcing stimulus after each trial.

The knob apparatus is essentially the same as that described in a previous publication (Rosenberg & Hall, 1958). In this study, however, reinforcing stimulus values were two- or three-valued rather than scalar quantities.

Subjects

*S*s were male Air Force basic trainees whose modal age was about 17 years. Trainees were assigned as a unit for use as *S*s in psychological research but were excluded form the experiment if their Armed Forces Qualification Test (AFQT) score was below the 50th percentile rank, which is roughly equivalent to a Stanford Binet IQ of 100 (Gragg & Douglass, 1956). *S*s were used in pairs, randomly assembled. Twenty pairs were used for Experiment 1, 60 pairs in Experiment 2, and 36 pairs in Experiment 3.

General Procedure[2]

*S*s in all experiments and in all experimental conditions were given the same instructions followed by 50 trials. Instructions were tape-recorded and heard by *S*s through earphones. An *E* pointed to and operated appropriate parts of the panel when they were mentioned in the instructions. Each *S* was told:

> The Air Force needs to know more about how well people can do certain tasks. The experiment in which you are to take part is a study of just one

[2] The author wishes to acknowledge the able assistance of Richard W. Hall and Stuart Dobson in the conduct of the experiments. Their dedicated efforts made it possible to complete this series of experiments.

such task. It is a simple task of turning a knob the right amount. Listen carefully to these instructions and the experimenter will point out things as they are mentioned. On the panel in front of you is a shield, and underneath the shield where you cannot see it, is a small knob. You may look under the shield now to locate the knob, but you are expected to work without seeing the knob. *Your job is to learn to turn the knob just the right amount. You will have a number of chances to do this. The green signal light will tell you when to start and stop. When the green light comes on, you begin to make your setting. As soon as you have made the setting you think is correct, take your hand away from the knob. When the green light goes off you must take your hand away whether you are finished or not. You must turn your knob each time the green light comes on. How much you turn it is what you have to learn. You will find that you have plenty of time without rushing. A few seconds after you have made your setting, you will be shown how well you did by the signal lights mounted just above the shield. Always turn your knob to the right as indicated by the sign tacked just over the shield. After each trial your knob is set back by the experimenter so that each time you make a setting you are starting from zero.* [long pause] Because it is important that you understand these instructions, I will repeat the most important parts. [Italicized portion repeated here.] Now we are ready to start. Do not touch your knob until the green signal light comes on. Now remove your earphones and watch for the green light to come on.

One complete trial consisted of three automatically timed intervals: (*a*) a response period of 8 sec. indicated to the *S* by a green signal light during which the knob could be turned; (*b*) a delay of 10–12 sec. during which *E* recorded the response magnitude of each *S* in the dyad, set these values into a computer, and read from the computer the values of reinforcement to deliver to each *S*; (*c*) the presentation of reinforcement by turning on the appropriate lights of *S*'s panels for 2 sec. After 2–3 sec. delay a new trial was started.

The presence of both *S*s in the same room did not appear to suggest to any of the *S*s that a behavioral link might exist between their tasks. A sizable number of *S*s made spontaneous comments about the task and "how it works" at the end of the session but none commented about the possibility of an interpersonal linkage. The nature and presentation of the instructions, the use of separate, completely enclosed booths for each *S*, the absence of an apparent mechanical linkage between the apparatus in the two booths all seemed to create the set for each *S* that he was being studied as an individual, albeit simultaneously with another person.

EXPERIMENT 1

Experimental Conditions

Each member of the dyad was assigned to one of five social reinforcement conditions. These five conditions differed from each other in that a different functional relationship was used to determine the S^r (reinforcement stimulus) to an *S* from the response magnitudes of both S_x and S_y.

To understand the derivation of the five reinforcement conditions, it is first necessary to define a variable Z. The five conditions were generated by assigning one of five values to the parameter, a_x *or* a_y, in the following equations

$$Z_x = a_x R_x + (1 - a_x) R_y$$
$$Z_y = (1 - a_y) R_x + a_y R_y$$

where $1 \geq a \geq 0$, and where R_x and R_y are the numbers of turns by each *S* on a particular trial. Note that the Z function is a simple weighted average of the two response magnitudes obtained on each trial and that the value of *a* is the proportion of variability in the Z values of an *S* that is attributable to his own response. If a_x were 1.00, for example, the situation would be reduced to individual trial-and-error learning for S_x. He would determine all of the variability of the Z values. All other values of *a* would be considered "social."

The numerical values of Z_x and Z_y were not the actual reinforcement stimuli delivered to S_x and S_y, respectively. The numerical value of Z on a trial determined which one of the following three S^r events were presented to an *S* on that particular trial:

1. Poor—Too Little	when $Z < 3.6$ turns
2. Good	when $Z = 4.0 \pm \leq .4$ turns
3. Poor—Too Much	when $Z > 4.4$ turns

This is essentially a coarse grouping of Z values, in which weighted average values below or above a certain range are Poor and with further specification as to direction. The same range of Z, $4.0 \pm \leq .4$ was used to transform both Z_x and Z_y.

When the parameters of both functions, Z_x and Z_y, have been specified numerically, and the relationship between Z values and S^r events stated, an "interpersonal reward system" is said to have been constructed.[3] In Experiment 1, five interpersonal reward systems were constructed by using the five *a* values: .55, .525, .50, .475, and .45, and im-

[3] Note that the term "reinforcement condition" refers to the operations by which an S^r to a single *S* is determined and that "interpersonal reward system" refers to the *combination* of "reinforcement conditions" for two or more persons. This terminology is used consistently throughout the paper.

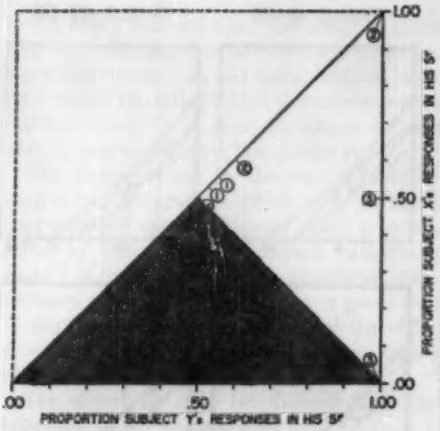

FIG. 1. Graphic summary of the cooperative reward systems used in the three experiments. The circled numbers represent the combinations of conditions for S_x and S_y in each experiment. The circled Digit 1 represents the combinations in Experiment 1, the Digit 2 in Experiment 2, and the Digit 3 in Experiment 3. The other half of the square would be a repetition of the same combinations with the conditions reversed for S_x and S_y.

posing the restriction that $a_x = a_y$ for each reward system. Thus, Experiment 1 was limited to reward systems in which each member of the dyad received the same *proportion* of *his own* response in his reinforcement. That is, when a_x was .525, for example, a_y was also .525. This should be carefully distinguished from a system in which both members receive the same reinforcement *values*, (e.g., a_x = .525, a_y = .475). Figure 1 is a graphic summary of the combinations of a_x and a_y chosen in Experiments 1, 2, and 3. The five combinations of a_x and a_y used in Experiment 1 are indicated by the figure 1 in a circle. The rationale for choosing these particular functions was derived from previous study. Previous data (Rosenberg, 1959) suggested that these five points represent the critical range of the continuum from those systems in which procurement and maintenance of reward develop for both *S*s to those systems in which continuous reward failed to develop for at least one of the two *S*s.

It should be noted that all five reward systems in Experiment 1 satisfy the definition of a cooperative situation on every trial. That is, there exist at least two outcomes for each *S* following a response, at least one is rewarding (Good) and at least one is not (Poor), reward to one *S* does not preclude its availability for the other, and, finally, the a values chosen determine that the outcomes to both *S*s are, in part or entirely, a function of the response of the other *S*. In fact, all combinations of a_x and a_y in Fig. 1, except where a_x or a_y equals 1.00, are cooperative interpersonal reward systems.

Results and Discussion

Figure 2 summarizes the essential findings of Experiment 1. Careful study of this figure will clarify the experimental conditions and the findings considerably. Results of subsequent experiments are presented in a similar manner.

The columns of small graphs represent the first two and last two successive blocks of five trials, as labeled across the top. These trial blocks seemed sufficient to demonstrate both (*a*) the direction of response changes during the 50 trials, and (*b*) the stability of the response values at the end of the experiment (comparisons of Trial Block 41–45 with 46–50). Each row of graphs represents one of the five reinforcement systems and is labeled in the first graph of each row. For example, S^r .525–.525 in the second row indicates that reinforcing stimuli were based upon Z_x = .525 R_x + .475 R_y and Z_y = .475 R_x + .525 R_y. The two coordinates, x and y, for each point plotted are based upon the averages of response magnitudes for five trials performed by S_x and S_y, respectively. A symbol shape (e.g., triangle, circle) is used consistently *within* each row to represent measurements from the same dyad, but similar symbol shapes, among different rows, refer to different dyads. A single symbol always represents responses made by the two members of a *single dyad*, never an average of several dyads.

Within a graph, the two pairs of parallel lines were derived from the equations for Z_x and Z_y. The area between a set of parallel lines is the Good range for either S_x or S_y. The area below and/or to the left of a set of parallel lines is the Poor—Too Little range and the area above and/or to the right is the Poor—Too Much range for an *S*. The combinations of these S^r events for each subarea are shown

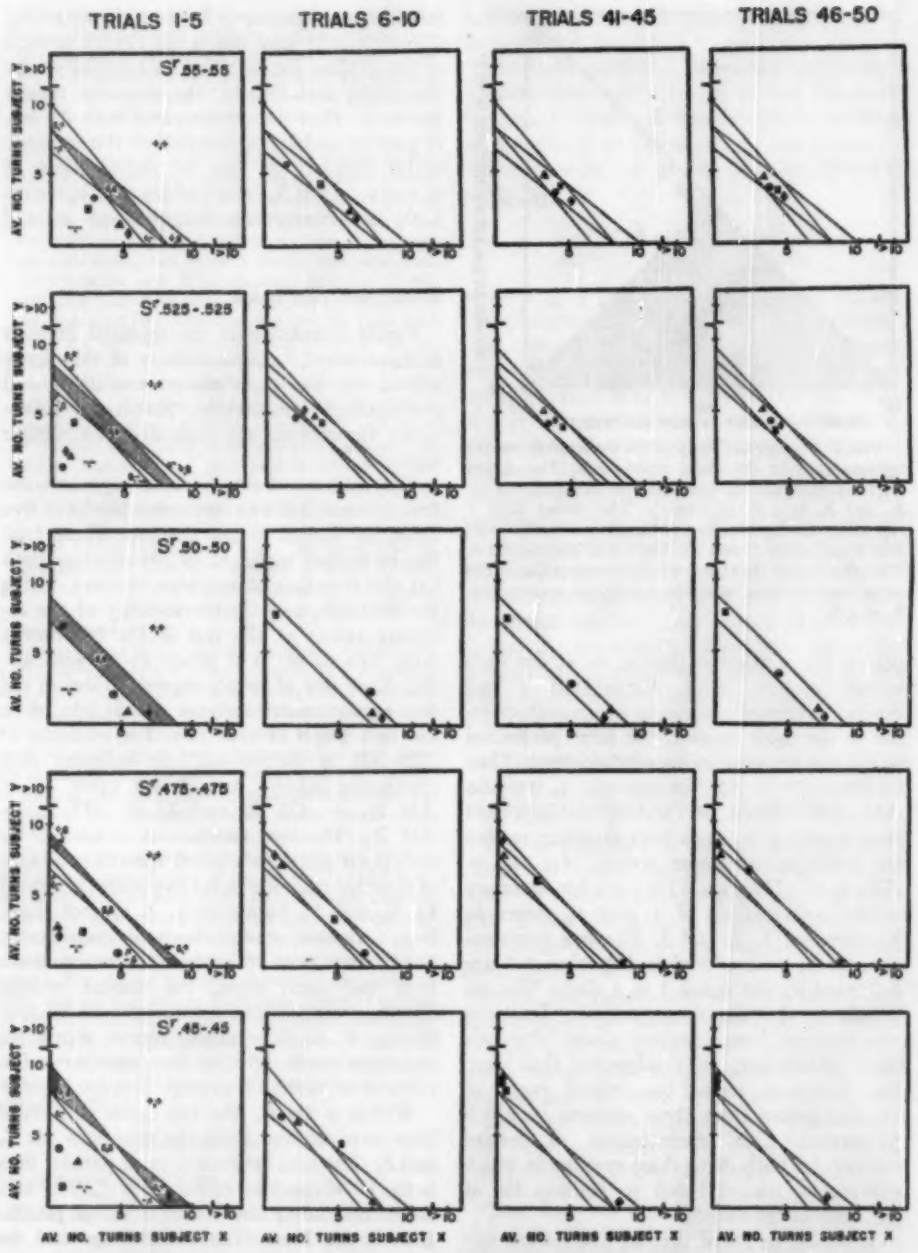

FIG. 2. Summary of the results of Experiment 1. The first two and last two successive blocks of five trials are shown as columns of small graphs and the five reinforcement systems as rows of small graphs. A symbol shape (e.g., triangle, circle) within each row represents the same dyad, but similar shapes among different rows refer to different dyads.

in the first graph of a row. For example, if the two response magnitudes were such as to fall anywhere in the (+, +) area, both *S*s were told Poor—Too Much; or, if the two responses fell into the (G, −) area, S_x was told Good and S_y was told Poor—Too Little; etc.

The shaded areas in the first column of graphs are the areas in which it is predicted that a pair of responses would stabilize over a series of trials and are termed "absorbing areas." The presence of two absorbing areas within a reward system indicates a prediction of response combinations in one *or* the other area, and not that there will be oscillation between these areas.

The absorbing areas are derived from the following simple assumptions:

1. An *S* will tend to decrease his response magnitude following a trial in which he is told Too Much and increase it if he is told Too Little. (This assumption fails to hold, of course, at the extremes, i.e., either when an *S* is giving an extremely small response and is told Too Much, or when he is giving a large response approaching the limit of his speed of turning and is told Too Little.)

2. If *S* is told Good the change in his response magnitude on the following trial is smaller than changes with the other two feedback stimuli (but not zero). (It is important to note that over a series of Good trials, these small response changes would not vary around a constant magnitude. Changes would either occur according to a "random walk" and/or tend to drift in the direction of least effort.)

3. The directions of the response changes by S_x and S_y following a given S^r are assumed to be additive and only the directions of response change are required to predict absorbing area(s). The rate of response changes is ignored in the summations to determine these areas. (Where two areas are predicted or where the predicted area is large, differential initial rates of response change probably determine which of the two areas or which portion of a single area will be the absorbing one. If one is interested in the proportion of trials spent in each area, response rates are also important.)

With these assumptions, a vector of response change may be drawn for each area within a graph. These vectors are all either vertical, horizontal, or at a 45° angle. A few examples are:

−, − ↗
G, − ↑
+, G ←
+, + ↙
G, G −

Having drawn vectors in all the subareas, one can predict the resultant movements of a combination of responses over a series of trials. In the first three rows in Fig. 2, the functions of Z generate a set of vectors converging on the (G, G) area from *all parts* of all other subareas. A response combination occurring in any subarea other than (G, G) would be followed by an S^r combination tending to change the responses toward the (G, G) area only. In the bottom two rows, however, vectors in two subareas, (G, +) and (+, G), do not point exclusively to the (G, G) area. That is, the S^r combination in a small portion of each of these two subareas is expected to change the response magnitude of one *S* but the change should still maintain the response combination in that subarea. Even if response combinations do occur in the (G, G) area during the process, these combinations are not stable over time (Assumption 2, above) unless all parts of all subareas lead to the (G, G) area. The ultimate absorbing areas in the bottom two rows are, therefore, (G, +) and (+, G) subareas and not the (G, G) area. The deeper shading within these two absorbing areas represents a prediction of responses to that part of the absorbing area, i.e., where one *S* is performing extremely small responses and the other *S* is performing responses somewhere within the range designated by the deeper shading. This area of response combinations is termed, for convenience, an "absorbing wall." The absorbing wall exists because there is no provision for "negative responses." In the case of this study, no turns of the knob in the opposite direction are possible. If this direction of turning *were* permitted, the prediction for these two reward systems would not be of the same areas but would be of areas with a similar consequence, i.e., a failure in the optimal development of cooperative action. The important point is not whether or not an absorbing wall exists, i.e., where one *S* can be told Too Much for very small response magnitudes and cannot make negative responses,

but rather that there are parametric values of interpersonal reward systems that lead to the breakdown of optimal cooperative action.

Other Z functions, outside the domain of the present study, may generate a set of vectors for which the resultant is not readily apparent and/or for which one would predict continuous movement from one subarea to another rather than the presence of absorbing area(s) or wall(s).[4]

Visual examination of the graphs shows striking accuracy of prediction for almost every dyad in every condition. The only reward system in which there appears to be any deviation from prediction is in S^r .475–.475 (fourth row) and this can probably be attributed to the relatively small absorbing area. Even here, however, there is no radical departure from prediction. The points are all in close proximity to the predicted areas. Comparison of Trial Block 41–45 with Trial Block 46–50 also shows, graphically, the high stability over time of response pairs in all conditions after 40 trials.

Two analyses were performed to establish statistical significance of these observations. The first tested the adequacy of the predictions over all conditions. The second analysis examined the possibility of differentially effective prediction among conditions. Both analyses were limited to data from the last 10 trials. For each of the 20 dyads, a count was made of those trials in which the response combinations fell outside the predicted area. This number was subtracted from the number of trials in which prediction was verified. These 20 difference values (theoretical range: $+10$ to -10) had a mean of $+5.20$ and σ of 5.60. The positive value for the mean indicated more hits than misses on the average. A t for differences against the null hypothesis that there were as many hits as misses yielded a value of 3.883 (*df* 19), significant beyond the .001 level.[5] The second analysis was a simple F test (between and within groups breakdown) in which the same difference scores were compared among the five systems. The null hypothesis was that each of the five conditions was equally predictable as measured by these difference scores. A qualitative comparison of the mean difference for each condition does suggest that S^r .475–.475 is somewhat less predictable than the others, but an F of 1.88 (*df* 4, 15) is not sufficiently large to reject the null hypothesis at the .05 level.

The statistical examinations of the data support the readily observable conclusion that availability of a cooperative situation does not assure optimal cooperative action, i.e., procurement of continuous positive reinforcement for both *S*s on every trial. If we view the five reward systems as points on a continuum, there appears to be a critical value below which *optimal* cooperative action fails to develop. In terms of Experiment 1, both *S*s receive continuous positive reinforcement when the reinforcement value to each is based upon 50% or more of his own response magnitude, i.e., when $a_x = a_y \geq .50$ in the two Z functions. Proportions less than 50% result in partial or no reinforcement to one *S*. Experiment 3 will demonstrate that $a_x = a_y$ is not a necessary condition for either of these effects and will provide a generalization for all reward systems depicted by Fig. 1.

Before additional empirical exploration, it is worth examining the features of these cooperative situations which covary with manipulations of the Z functions. It is obvious from Fig. 2, for example, that the relative sizes of the subareas vary from row to row. And in S^r .50–.50 four subareas are completely absent. However, there is little reason to interpret the results in terms of available areas or their sizes relative to each other. The selection of systems above and below the middle row was such as to control for subarea sizes. Thus, for example, S^r .55–.55 and S^r .45–.45 have comparable subareas and they are of the same size.

It is possible to suggest other interpretations

[4] Many of the two-person game matrices of current interest to probability learning theorists and to social psychologists appear to result in asymptotic response probabilities in a cell or "area" of less than 1.0, i.e., continuous movement from one cell to another. More quantitatively stated assumptions are required to predict these response probabilities for each area in such interpersonal reward systems (Atkinson & Suppes, 1958; Estes, 1957). It is interesting to note that response *sequences* in these systems rarely receive analytic treatment. The present paper is an attempt in this direction.

[5] This is a rather conservative estimate of the accuracy of the prediction since it fails to credit occurrences in proximal areas, i.e., "near misses" and since predicted areas are very small relative to total area.

of the results in reward systems that fail to elicit optimal action. It may be that the absorbing areas are valid because of the improbability that *S*s ever encounter the (G, G) outcome. If they did, this interpretation would predict that their responses would stabilize in (G, G) rather than in (G, +) or (+, G). The data fail to support this interpretation. In S^r .45–.45 and S^r .475–.475 there is an average of 7.0 and 19.8 (G, G) outcomes in the first 40 trials. This fact is discernible to some extent in Fig. 2 for Trial Blocks 6–10. In addition, results of a previous study (Rosenberg, 1959) demonstrated very clearly that even if *S*s are first individually trained to a point where they are both receiving continuous reinforcement, i.e., placed in the (G, G) area, the change to a cooperative condition results in gradual deterioration of reinforcement if the relative response weighting is less than S^r .50–.50.

Finally, one may explain stability of a nonoptimal absorbing wall in terms of partial reinforcement for one *S* and continuous reinforcement for the second *S*. Since the (G, G) area is adjacent to the (G, +) and (+, G) areas, the (G, G) outcome may occur with sufficient frequency to reinforce S_x or S_y (as the case may be) for very small response magnitudes, and thereby produce relatively stable behavior. This explanation was tested in Experiment 2 by studying more extreme conditions where partial reinforcement was less likely to occur, i.e., where the (G, G) area was not adjacent to (G, +) or (+, G) areas. This extension was combined with an analysis of a second parameter of the reinforcing stimulus, i.e., its informational ambiguity.

EXPERIMENT 2

In essence, the results of Experiment 1 demonstrated that *S*s who are placed in a reward system in which relatively few outcomes may be misleading fail to gravitate toward optimal cooperative behavior in spite of the presence of an optimal outcome and in spite of the occurrence of such an optimal outcome during the learning sequence. This finding suggests that a cooperative situation be constructed using an unordered set of reinforcing events, and hence, no "misleading" outcomes. Such unordered S^r events are less explicit either because (*a*) they are only qualitatively discriminable to *S*, although outcomes do implicitly covary with the quantitative combinations of R's or (*b*) unrewarding outcomes actually contain no information about what kind of response change is needed to bring about the rewarding outcome. The reduction of explicit information by either alternative may retard learning in cooperative situations where explicit information would have resulted in rapid learning. On the other hand, an unordered set of outcomes may produce behavior that yields a larger proportion of rewarding outcomes in cooperative situations where more explicit information about Z values would be misleading. Thus, three sets of reinforcing events varying in informational content about Z were compared in Experiment 2. Information about the direction of nonreinforcement has been studied extensively in individual learning (Stevens, 1951, pp. 1267–1270), but has received scant attention as a parameter in interpersonal learning. Finally, a wider range of *a* values was studied in Experiment 2 in order to assess the limits of generality of results in Experiment 1.

Experimental Conditions

Each dyad was assigned to one of the five reward systems indicated in Fig. 1 by the number 2 in a circle, and to one of the three sets of reinforcing events. Thus, there were 15 experimental conditions (5 × 3) with four dyads in each condition.

In terms of the equations

$$Z_x = a_x R_x + (1 - a_x) R_y$$
$$Z_y = (1 - a_y) R_x + a_y R_y$$

a_x equalled a_y within a dyad. As in Experiment 1, therefore, S_x received the same proportion of R_x as did S_y of R_y in each dyad. The five values of *a* were 1.00, .60, .50, .40, .00. It might be noted that the first combination of values of *a* (1.00, 1.00) is at an extreme which is completely asocial since neither *S* depends upon the other for his reinforcement. It is the only combination in Experiment 2 which fails to fulfill the definition of a cooperative situation.

The three sets of reinforcement events varying in information about Z were:

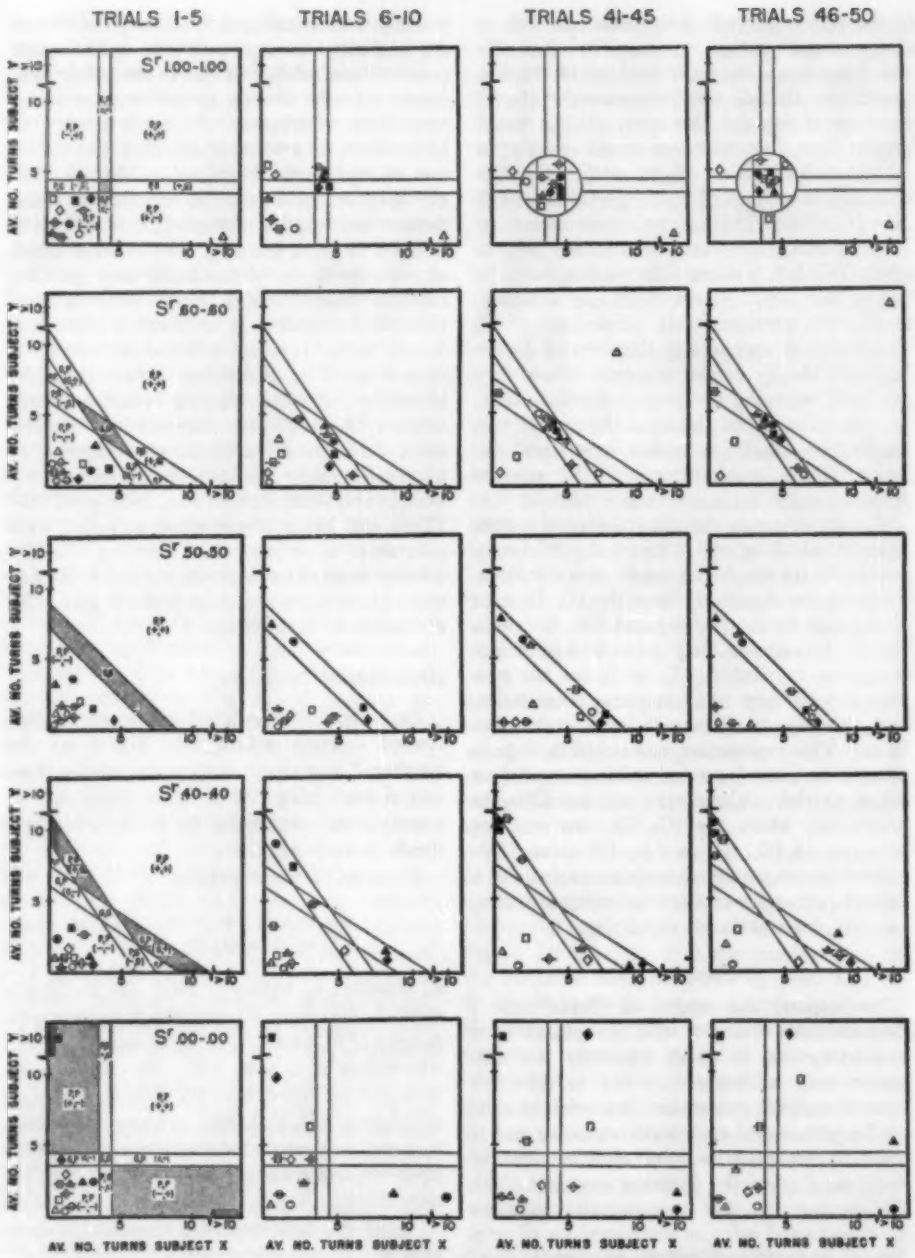

FIG. 3. Summary of the results of Experiment 2 (see explanation of Fig. 2). The filled symbol shapes refer to dyads given explicit information about plus or minus values shown in each subarea of a graph; the open symbols with horizontal bar refer to dyads given ambiguous information about the plus or minus values; the open symbols represent dyads given only "poor" or "good" as outcomes.

1. Poor—Too Much; Good; Poor—Too Little (MGL)
2. Poor; Good; Poor (PGP)
3. Poor; Good (PG)

The first set is the same as that used in Experiment 1. In the second set, Too Much and Too Little were left off, but which of the two Poor outcome lights was given on a trial depended consistently on whether the Z value was above or below the Good range. In the third type, all Z values above and below the Good range were called Poor, and hence only two outcome lights were used, one for Good and one for Poor.

Results and Discussion

Figure 3 is a graphic summary of the results. Format is similar to that employed in Fig. 2. Additional notations were required to distinguish the three sets of reinforcement stimuli. The solid black symbols represent the MGL set (used exclusively in Experiment 1), the symbols with horizontal bar represent the PGP set, and the open symbols represent the PG set of two stimuli. Use of the same symbols within a row indicates successive measurements from the same dyad while similar symbols in different rows refer to different dyads.

Also, each subarea on the graph is labeled with two pairs of coordinates. Coordinates in parenthesis refer to the explicit or implicit values associated with the MGL and PGP conditions, respectively. Coordinates outside of parenthesis refer to a pair of outcomes in the PG condition. The shaded areas were derived from the same assumptions stated in Experiment 1, although the validity of these assumptions for the PGP set and certainly for the PG set is questionable.

An analysis was first performed to ascertain the differential validity of these assumptions on the last 10 trials for all 15 conditions. A factorial design was employed containing two criteria of classification: (*a*) five reward systems and (*b*) the three sets of reinforcement events varying in informational content. The dependent variable for each dyad was simply the number of trials between 41 and 50 in which the response combination occurred in the predicted area minus the number of occurrences in all other areas. The analysis is summarized in Table 1 and indicates a significant *F* only among the three reinforcement sets varying in information. The graphs in Fig. 3 illustrate very clearly the nature of these differences. The assumptions for deriving the shaded area are least valid for the PG condition. There are almost no dyads in this condition whose behavior conform to prediction. The assmptions are more valid for the PGP conditions but still fall short of the very consistent findings for MGL condition. Deviation from prediction does occur in the MGL condition also, particularly near the extreme where the *a* values are small, i.e., where cooperative action is not predicted to develop optimally. In S^r .00–.00, where it was predicted that neither *S* would receive reinforcement, the four dyads conform to prediction in the early trials, e.g., 6–10, but by the end of the experimental session two dyads have moved out of the predicted areas. Also in S^r .40–.40, one dyad changed responses so as to move out of the predicted area during the last five trials. These response changes out of predicted areas, interestingly enough, do not result in more effective cooperative action, at least during the number of trials studied. These findings for small *a* values lend support to the hypothesis suggested in discussion of Experiment 1, i.e., that a nonoptimal absorbing area may depend for its long-term stability on partial reinforcement of the small response magnitudes by one *S*. A significant interaction of reward system with information content, i.e., 1 × 2 in Table 1, could support this interpretation more fully. The absence of interaction fails to refute, however, the generality of the assumptions used to derive predicted subareas. More *S*s and more trials may be required to refute the validity of the assumptions for extreme

TABLE 1

ANALYSIS OF VARIANCE OF THE LAST 10 TRIALS OF EXPERIMENT 2 USING THE DIFFERENCE SCORES (FREQUENCY OF PREDICTED AREAS MINUS FREQUENCY OF OTHER AREAS) AS THE DEPENDENT VARIABLE

Source	*df*	Mean Square	*F*
Reward System	4	20.83	.70
Reinforcement Sets Varying in information	2	856.06	28.96*
1 × 2	8	27.73	.94
Within Cells	45	29.62	(error term)

* $P < .001$.

reward systems involving the set of outcomes, MGL.

Accurate prediction for the PGP and PG conditions obviously cannot be completely derived from these assumptions. The PGP condition results in behavior that conforms, in part, to the predictions. Apparently the early trials provide a sufficiently consistent relationship between changes in response magnitude and presentation of the two Poor outcomes so that *S*s learn to "impose" an appropriate ordering of these two Poor outcomes, i.e., discriminate between them. The particular reward system may be such as to continue to reinforce this discrimination of the two Poor outcomes. However, in other reward systems (bottom two rows) the decrement in reward frequency may result in the extinction of this implicit ordering behavior. Whether, at this point, the *S* reverses the values, i.e., reverses his previous discrimination, or simply lumps the two Poor outcomes, i.e., the discrimination is extinguished, or is inconsistent is difficult to infer from these data. In any case, however, the frequency of reward does not increase appreciably over time in the conditions represented by the bottom two rows. There is, simply, a lower frequency of the very large or very small response magnitudes that are found in the MGL condition.

The validity of the predicted areas for the PG condition is nil. This finding is not surprising since the assumptions from which these areas are derived require informational components in the reinforcement not present in the PG condition. The *S* is given no information about response magnitude, and hence there is no reason to assume either a consistent increase or a consistent decrease in response magnitude following a Poor outcome, i.e., the first assumption for an MGL condition would be discarded. One could retain the two remaining assumptions, however, which are essentially that an *S* will tend to make a larger change in response magnitude (sometimes increasing it and sometimes decreasing it) following a Poor outcome than following a Good outcome and that the directions of response changes by the two *S*s are additive. With these assumptions, it is possible to derive a number of interesting statements that contrast with those made for the other two types of conditions. First, one would predict that in the *cooperative* reward systems in this study, the most stable subarea, if any, would be the (G, G) area. That is, if a response combination falls into the (P, G) areas, S_x will make large response changes. Similarly, if the combination is (G, P), S_y will make large response changes. A (P, P) outcome will yield large response changes by both *S*s. The sizes of these areas relative to the (G, G) area should, therefore, be *one* determinant in the frequency of reward for *both* *S*s. Hence, S^r .50–.50 might be expected to produce more Good outcomes than adjacent conditions since it has the largest (G, G) area and no (P, G) or (G, P) areas. It is the one condition in which both *S*s make large response changes or neither does.[6] This seems to be the effect for at least two of the four dyads in this condition. By contrast, neither the adjacent conditions S^r .60–.60 nor S^r .40–.40 reward systems produce nearly such a high proportion of reward for both *S*s.

The relevance of the Z functions themselves in the PG condition is also derived from this reasoning. We have assumed that (P, G) and (G, P) are unstable to some extent because either S_x or S_y, respectively, makes large response changes. The extent to which these outcomes *are* unstable depends in turn on the pair of Z functions. Thus, if the outcome is (P, G) and the response of S_x received a large weighting in the reinforcement of S_y, it would be expected that S_y would soon fail to receive positive reward. This in turn increases the size of response change of S_y. On the other hand, Z functions that make one or both *S*s relatively independent of the other would produce a greater number of Good outcomes for one or both *S*. This is discernible in the extreme condition S^r 1.00–1.00 where both *S*s are independent of each other. Nevertheless, S^r .50–.50, which appears to clearly divide optimal from nonoptimal reward systems in the MGL condition, is not the critical point in the PG condition.

[6] It is interesting to note that the S^r .50–.50 is the only reward system which satisfies some definitions of a cooperative situation, viz., when one receives reward, all do, when one fails to receive reward, no one does (Lindzey, 1954, p. 810). One may speculate about the possibility that this restrictive definition evolved because cooperative action develops best here with equivocal outcomes and that the tasks typically used in the past for cooperative studies do contain more equivocal reinforcing events than those in Experiment 1.

It would be interesting to pretrain both *Ss* individually so that their response combinations fell into the (G, G) area and then change them to a cooperative situation. Because of the response drift, and on the basis of this experiment, it is quite probable that a (G, P), (P, G), or (P, P) condition would soon occur and that return to (G, G) would then be as difficult for such a dyad as for one which had received no preliminary individual training.

In summary, then, using a PG set of outcomes, one would probably predict that no subarea is as stable as those found in more informative outcomes. In general, one would predict that the (G, G) area is most stable, the (P, G) or (G, P) areas next most stable, and (P, P) areas least stable. The frequency of occurrence of responses in any of these areas depend upon (*a*) the relative sizes of these areas and (*b*) the degree of independence between persons as determined by the pair of Z functions.

Quantitative assumptions are required to make more exact predictions about outcome stability as a function of these experimental parameters. Qualitatively, it seems surprising, nevertheless, that the wide variations in the Z functions in this study (short of complete autonomy) result in such an ineffective procurement of positive reinforcement for both *Ss* over all 50 trials in the PG condition. There is some reason to think that this poor performance is, in part, a function of the type of *S* used in this study. It was observed that in many dyads both members failed to demonstrate much exploratory behavior over the entire 50 trials. They failed to increase their response magnitudes sufficiently to get out of

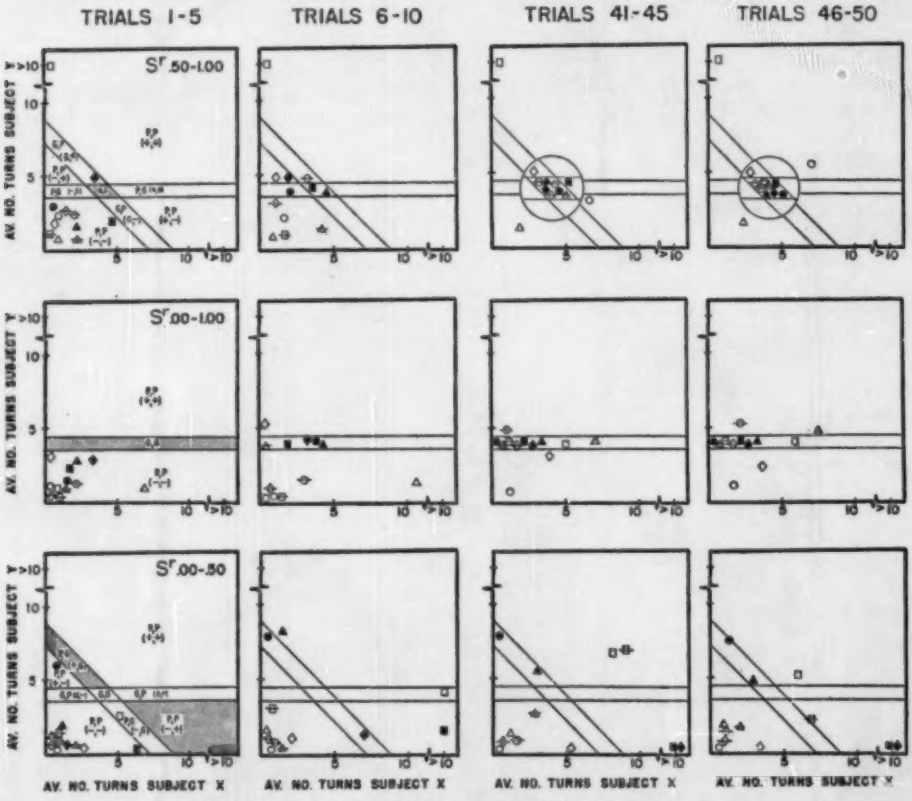

FIG. 4. Summary of the results in Experiment 3 (see explanation for Fig. 3).

the lower left (P, P) area. This depressed activity may be inherent in the nature of the sample of *S*s. They were all persons who had just entered Air Force service and hence were in a very strange and stressful environment generally. It is unknown whether other, more relaxed *S*s may have simply elevated the total frequency of reward over all reward systems, or whether the effects of different reward systems would have been more clearly distinguishable using the PG outcomes with the more "relaxed" *S*s.

EXPERIMENT 3

The purpose of the final experiment is best understood by reference to Fig. 1, the graphic representation of the combinations of Z functions for each dyad. The results of the first two experiments indicate that a predicted subarea for the most explicit form of reinforcement stimulus, MGL, is unrewarding to one or both *S*s when a reward system is imposed in which both receive as feedback less than half of their own response variance in the S^r and when $a_x = a_y$. This is the lower half of the hypotenuse of the triangle in Fig. 1. By extrapolation, it may be demonstrated that these two restrictions are not necessary conditions for predicting an absorbing area containing an unrewarding outcome for one *S*. In fact, a nonoptimal absorbing area will be predicted even though one *S* may receive a *majority* of his response in his reinforcement as long as

$$a_x + a_y < 1.00$$

This equation defines the shaded area in Fig. 1. When $a_x \neq a_y$ the effect is to increase the probability of unrewarding outcomes to the *S* assigned to the smaller *a* value and decrease the probability of unrewarding outcomes to the *S* assigned the larger *a* value.

The purpose of this experiment is to demonstrate these predictions empirically by selecting three reward systems not on the hypotenuse of the triangle where

$$a_x + a_y > 1.00$$
$$a_x + a_y = 1.00$$
$$a_x + a_y < 1.00$$

Experimental Conditions

Each dyad was assigned to one of the three degrees of S^r information, i.e., MGL, PGP, PG, and to one of the three reward systems shown in Fig. 1 by the number 3 in a circle. This made a 3 × 3 design with four dyads in each cell.

The sum of values for a_x and a_y were 1.50, 1.00, and .50. The *a* values were, respectively, (*a*) .50, 1.00; (*b*) .00, 1.00; (*c*) .00, 50.

TABLE 2

ANALYSIS OF VARIANCE OF THE LAST 10 TRIALS OF EXPERIMENT 3 USING THE DIFFERENCE SCORES (FREQUENCY OF PREDICTED AREAS MINUS FREQUENCY OF OTHER AREAS) AS THE DEPENDENT VARIABLE

Source	*df*	Mean Square	*F*
Reward System	2	85.77	4.92*
Reinforcement Sets Varying in information	2	341.77	15.60***
1 × 2	4	95.45	5.47**
Within Cells	27	17.44	(error term)

* $P < .05$.
** $P < .01$.
*** $P < .001$.

Results and Discussion

Figure 4 is a graphic summary of the results. The same notation was used here as in previous figures.

Qualitative inspection supports the validity of predictions for the MGL condition. The two other types of S^r information show less tendency to produce response pairs in the predicted areas, and again, the PG condition is least predictive. Note also that the PG condition is in general somewhat less sensitive to the different reward systems than are the other two informational conditions.

An analysis of variance of the results for the last 10 trials is summarized in Table 2 and corroborates the significant differences among the three reinforcement types. Significant effects attributable to different reward systems and to interaction were also obtained. Figure 4 suggests the nature of these effects. The PGP and, to some extent, the PG conditions show better predictability when (G, G) is the predicted outcome than when (P, G) or (G, P) is predicted. This result is in line with the discussion of these conditions in Experiment 2.

GENERAL DISCUSSION

There appears to be a growing interest in the simplification and increased control of laboratory situations employed to study interpersonal phenomena. This trend stems, in

part, from the analyses of interpersonal behavior based upon an S-R system. (Lindzey, 1954, pp. 77–87; Logan, 1955; Miller & Dollard, 1941; Sears, 1951.) These analyses usually include concepts and principles of learning, and the empirical studies almost invariably employ laboratory situations that are highly controlled with regard to the stimulus and response elements of an interpersonal event (e.g., Miller & Dollard, 1941; Rosenberg & Hall, 1958; Rosenberg, 1959; Sidowski, Wyckoff, & Tabory, 1956). Some writers, whether they apply an S-R system or not, employ these highly controlled situations for the application and development of rather formal theoretical systems, e.g., mathematical models of interpersonal events (Atkinson & Suppes, 1958; Estes, 1957; Hays & Bush, 1954).

The present study is concerned with the problem of organizing a set of interpersonal variables on a descriptive S-R level and in refining experimental techniques rather than in quantitative theory construction. The descriptive approach, however, is not aimed at the proliferation of experimental variables without regard to some parsimonious grouping of experimental operations. The present study, for example, would appear to provide generalization for a number of Z functions in addition to the simple additive ones used here. It would appear to matter very little, for example, whether the functions for S_x and/or S_y were curvilinear rather than linear. If one could determine unequivocably that absorbing area(s) existed for a pair of Z functions (and this is possible for a variety of nonlinear functions), it would matter little that these functions were nonlinear.

There are combinations of functions, however, for which no absorbing area(s) can be predicted from the assumptions in this paper. For example, if one used the function $Z = aR_y - bR_y$ for one S and $Z = a'R_x + b'R_y$ for the second S, one would be able to predict from the assumptions used in this study only that R_x and R_y would change continually from trial to trial so as to "orbit" around the (G, G) area. There is no way of predicting from present assumptions whether the orbit would become smaller and hence converge on the (G, G) area, become larger, change shape, etc. Experimental results using this interesting pair of general functions would presumably permit some generalization to a variety of similar combinations of functions. It is interesting to note, parenthetically, that the $Z = aR_x - bR_y$ function for one S in combination with a function describing individual discrimination learning by a second S resembles the imitation paradigm of Miller and Dollard (1941) and that an absorbing area does exist for this reward system in the (G, G) area.

Another interesting set of function pairs is that in which no (G, G) area exists. That is, the Z functions do not intersect anywhere on the graph. This specification would define a competitive task, if only one trial were permitted. Over a series of trials, however, such a combination of functions would fall within the definition of a cooperation. It would not be surprising, however, to find dyads unable to develop optimal cooperative action nor even to procure an optimal sum of individual reward frequencies although the situation permitted the occurrence of a rewarding outcome to one or the other S on each trial. Studies that combine interests in learning and in game matrices usually employ examples of this set of function pairs.

A second class of variables subject to empirical analysis and generally independent of the Z functions, are those termed "awareness". While these variables have received considerable prominence in social psychology (some social psychologists would reject a situation as interpersonal if Ss were not "aware" of social linkages), there has been little systematic attempt to formulate a parametric approach for this class of operations. When "awareness" is manipulated by E directly, it usually takes on only two values—an S is or is not told something. (Neimark & Rosenberg, 1959; Sidowski, 1957) Some experimental tasks, e.g., face-to-face, are so constructed that extensive manipulation of awareness is not possible. Other tasks such as the one in this study permit the E to assign many "values" to awareness. These values may range from telling each S nothing about the interpersonal linkage during the initial instructions, to a complete description for both Ss of the functions by which their responses are to be combined. Note also that the task in this study permits the E the choice of delivering to each S rewards of which the other S may or may not be made aware

from trial to trial, and thereby extended the concept beyond the instructional differences.

The implications of the present study would appear to extend beyond that of a knob-turning task. Pilot replications of aspects of this study by the author using other trial-and-error tasks, e.g., time estimation, size estimation, suggest ready generalization of results to these tasks. One would also expect, for example, some generalization to the autokinetic response. While it is premature to outline implications of the present results to life situations, it seems likely that the interpersonal reward systems imposed by laboratory situations can be coordinated with systems of reward imposed upon a dyad by some natural or mechanical environments—and perhaps can be eventually coordinated to reward systems imposed by the behavior repertoires of the interactors.

SUMMARY

Reinforcement parameters for a dyadic cooperative situation were defined and varied in a series of three experiments. A cooperative situation was defined as one where (*a*) there exist at least two outcomes for each *S*, (*b*) one of which is a rewarding state of affairs, (*c*) reward delivered to one person does not preclude its availability for the other, and (*d*) the outcomes to *both* persons are, in part or entirely, a function of the behavior of the other person. The task was the same in all three experiments. Each *S* was required to perform the simple motor response of turning a small concealed knob in a series of 50 trials. A reinforcing stimulus was delivered to each *S* separately after each trial. All conditions either satisfied the definition of a cooperative situation or were a social (wherein reinforcement was contingent upon individual behavior alone). The degree to which optimal cooperative behavior, i.e., procurement of maximum reinforcement by both *S*s developed in these situations, was studied as a dependent variable.

Experiment 1 was composed of five cooperative reward systems. These systems varied in the extent to which the response magnitude of one *S* of a dyad determined the reinforcement stimulus for the other *S* and vice versa. That is, the reinforcement stimulus, Good, was delivered to an *S* on trials when the Z value in the following equation fell within a predetermined range.

$$Z = aR_x + (1 - a)R_y$$

where R_x and R_y are response magnitudes of the two *S*s and where *a* is assigned a value between zero and one for each *S*. Predictions were made concerning the degree to which cooperative action would develop as a function of these relative values of *a* for each *S*. It was predicted that in three reward systems, where *both S*s' reinforcement stimulus was contingent upon a fraction equal to or greater than .50 of his own response magnitude, optimal cooperative behavior would develop quickly and would be maintained consistently. In the two remaining reward systems, however, where both *S*s' reinforcement stimulus was contingent upon less than .50 of his own response magnitude, optimal cooperative behavior was not predicted. The results clearly substantiated these predictions.

Experiment 2 extended the range of relative response weightings to its full range, i.e., *a* values from zero to one were used for different sets of dyads. This extended range tested the generality of the predictive assumptions used in Experiment 1. Predictions were substantiated for this extended range of weightings. The weighting, .50, was again demonstrated to be the critical point, i.e., when reinforcements to both *S*s were weighted by less than .50 of their own responses, optimal cooperative action failed to develop.

Informational content concerning the size of Z was varied in Experiment 2. It was found that predictions about *S*s' behavior were less valid as the information to *S* about Z became more equivocal and that the total frequency of positive reinforcement to either *S* dropped considerably under these conditions of reduced information.

Experiment 3 investigated further extensions of the response weighting parameter so as to demonstrate the generality of the predictive assumptions for all *combinations* of weightings. Experiments 1 and 2 were limited to combinations of weightings where the response of one *S* was weighted the same in his reinforcement stimulus as that of the second *S* in *his* reinforcements. Experiment 3 demonstrated when

one *S*'s response weighting in his reinforcement is different from the second *S*'s response weighting, optimal cooperative action fails to develop where the sum of these two weights is less than 1.00, and hence such failures are not limited to reward systems where *both Ss*' responses are weighted less than .50.

These experiments support the conclusion that the combination of response weightings in the reinforcements to each *S* is an important parameter in the development of cooperative action. The amount of information concomitant to the presentation of reinforcement to each *S* also appears to play an important role in the development of cooperative action and in the frequency of reinforcement procured by both *Ss* over a series of trials.

The status of the present study as a contribution to a descriptive system of interpersonal behavior was briefly discussed. Other experimental parameters were suggested and the implications of these parameters for analysis of interpersonal situations were sketched.

REFERENCES

ATKINSON, R. C., & SUPPES, P. An analysis of two-person game situations in terms of statistical learning theory. *J. exp. Psychol.*, 1958, **55**, 369–378.

DEUTSCH, M. A theory of cooperation and competition. *Hum. Relat.*, 1949, **2**, 129–151.

ESTES, W. K. Of models and men. *Amer. Psychologist*, 1957, **12**, 609–617.

GRAGG, D. B., & DOUGLASS, H. J. Conversion tables for selected Airman Classification Battery scores and comparable scores on other selected service and civilian tests. *USAF Pers. Train. Res. Cen. Developm. Rep.*, 1956, No. AFPTRC-TN-56-89.

HALL, R. L. Group performance under feedback that confounds responses of group members. *Sociometry*, 1957, **20**, 297–305.

HAYS, D. G., & BUSH, R. A study of group action. *Amer. Sociol. Rev.*, 1954, **19**, 693–701.

LINDZEY, G. (Ed.) *Handbook of social psychology*. Vols. 1 & 2. Cambridge: Addison-Wesley, 1954.

LOGAN, F. A. *Behavior theory and social science*. New Haven: Yale Univer. Press, 1955.

MILLER, N. E., & DOLLARD, J. *Social learning and imitation*. New Haven: Yale Univer. Press, 1941.

MINTZ, A. Non-adaptive group behavior. *J. abnorm. soc. Psychol.*, 1951, **46**, 150–159.

NEIMARK, EDITH D., & ROSENBERG, S. The effect of "social" discriminative cues on probability learning. *J. exp. Psychol.*, 1959, **58**, 302–311.

ROSENBERG, S. The maintenance of a learned response in controlled interpersonal conditions. *Sociometry*, 1959, **22**, 124–138.

ROSENBERG, S., & HALL, R. L. The effects of different social feedback conditions upon performance in dyadic teams. *J. abnorm. soc. Psychol.*, 1958, **57**, 271–277.

SEARS, R. R. A theoretical framework for personality and social behavior. *Amer. Psychologist*, 1951, **6**, 476–483.

SIDOWSKI, J. B., WYCKOFF, L. B., & TABORY, L. The influence of reinforcement and punishment in a minimal social situation. *J. abnorm. soc. Psychol.*, 1956, **52**, 115–119.

SIDOWSKI, J. B. Reward and punishment in a minimal social situation. *J. exp. Psychol.*, 1957, **54**, 318–326.

STEVENS, S. S. (Ed.) *Handbook of experimental psychology*. New York: Wiley, 1951.

TINBERGEN, N. *Social behavior in animals*. New York: Wiley, 1953.

(Received January 29, 1959)

Journal of Abnormal and Social Psychology
Vol. 60, No. 3, 1960

ACHIEVEMENT RELATED CHARACTERISTICS OF MANIC-DEPRESSIVES[1]

JOSEPH BECKER[2]
Duke University

THE general purpose of the present study was to attempt a more objective and rigorous testing of hypotheses about manic-depressive character derived primarily from the clinical research of Cohen, Baker, Cohen, Fromm-Reichmann, and Weigert (1954).[3] Their descriptions of manic-depressive character were reformulated in terms of personality types which have emerged from experimental investigations of human motivation. The Achievement concepts of McClelland and his associates (McClelland, Atkinson, Clark, & Lowell, 1953; De Charms, Morrison, Reitman, & McClelland, 1955) provided a particularly appropriate framework for examining certain aspects of manic-depressive character. The formulations of Cohen et al. attached considerable significance to behaviors and attitudes related to achievement and conformity, which are also of central importance to McClelland's concepts.

According to these writers (Cohen et al., 1954, pp. 113–119), early parental indulgence was abruptly supplanted by approval conditioned upon the fulfillment of parental demands for rigidly conforming behavior and high achievement. This change in the nature of parental demands resulted in rigid, impersonal standards being foisted upon the premanic-depressive child by his parents in order to use him as an instrument for raising the family's social prestige. These standards of achievement and conformity were determined by the parents' (usually the mother's) conception of what the neighbors would think proper and appeared to be severely enforced.

In adulthood (Cohen et al., 1954, pp. 119–121), the manic-depressive was caught in the dilemma of strongly inculcated achievement values based on the need to placate others, and and fear of alienating others through his success. Achievement was valued primarily as a means toward the end of gaining others' approval. Excellence of performance or independent activity as satisfying in itself was subordinated to the belief that success was attainable only through manipulating others, not through one's own efforts. So excessive is the dependence of the manic-depressive on the approval of others that he will abandon even his rigidly conventional opinions in favor of the perceived values of the needed other, particularly if the other is seen as an authority figure.

McClelland and his associates identified two relatively independent types of achievement orientation. Need Achievement is said to characterize persons who are concerned with living up to an internalized standard of excellence. Their satisfaction is in doing a job well rather than in the enjoyment of the end product. As children, high need Achievers were subjected to more early independence demands and had fewer restrictions imposed upon them than did low need Achievers (Winterbottom, 1953). In contrast, value Achievement characterizes individuals who have responded to excessive parental verbal emphasis on achievement striving, by a conscious valuing of achievement. DeCharms et al. (1955) in speculating about the origins of value Achievement state that persons "who describe themselves as ambitious and achievant may do so for defensive reasons; they have perhaps been under some authoritarian pressure from their parents to be ambitious and the resultant motive which has originated in external sources shows itself primarily as a fear of being unsuccessful" (p. 419). Whereas high need

[1] This paper is based upon a doctoral dissertation submitted to Duke University. The writer wishes to express his appreciation to Charles D. Spielberger, the dissertation chairman, and Joseph B. Parker, Chief of Psychiatry VA Hospital, Durham, North Carolina, for their generous assistance.

[2] Now at Umstead State Hospital, Butner, North Carolina.

[3] This investigation was based on a research project which studied 12 manic-depressives who were being treated by intensive psychoanalytic therapy. The research group was comprised of staff members of the Washington School of Psychiatry who met in bimonthly, three-hour seminar sessions for approximately four years.

Achievement has been shown to be positively related to better performance on tasks involving "internalized standards of excellence" and negatively related to conformity and authoritarianism (Brown, 1953; McClelland et al., 1953; Wendt, 1955), high value Achievement is unrelated to performance involving internalized standards of excellence and is positively related to conformity behavior and authoritarian attitudes (DeCharms et al., 1955).

According to the conceptualization of Cohen et al. (1954), manic-depressives and high value Achievers seem to share a common defensive reaction pattern consisting of outwardly adopting the prevailing values of others in order to placate and win needed approval from them. Dependency on others for guidance and approval is manifested in the behavior of both manic-depressives and high value Achievers through conventional attitudes, stereotyped social interactions, and conformity. However, Cohen et al. have also asserted that premanic-depressive children frequently assumed early responsibility, and as adults often became financially successful. Such behavior would not be inconsistent with what might be expected of McClelland's high need Achievers. Furthermore, Arieti (1957) has implied that manic-depressives are like Riesman's inner-directed character type (Riesman, Glazer, & Denny, 1950), a type with which McClelland (1955) has also identified the high need Achiever.

Since phenotypical characteristics and the socialization process of manic-depressives seemed to resemble those of high value Achievers more closely than those of high need Achievers, the hypotheses of this study predicted that manic-depressives would perform like high value Achievers, as compared to a nonpsychiatric control group, on a number of tasks which differentiate need and value Achievement performance.

The following predictions were made:

Hypothesis 1: Manic-depressives describe themselves as more ambitious and achievant than the controls.

Hypothesis 2: Manic-depressives subscribe more strongly to authoritarian ideology than the controls. Both high authoritarians and manic-depressives have been reported to share such characteristics as intolerance of ambiguity, social imperceptiveness, stereotypy and rigidity in behavior and attitudes, and strong conformity tendancies (Titus & Hollander, 1957; Cohen et al., 1954).

Hypothesis 3: Manic-depressives profess a more traditional ideology with regard to family roles and statuses than the controls.

Hypothesis 4: Manic-depressives are not significantly different from the controls in strength of need Achievement phantasy.

Hypothesis 5: Manic-depressives are not significantly different from the controls in their achievement on performance tasks under either task or achievement oriented conditions.

If, however, the manic-depressives more closely resemble high need Achievers, then on the operational measures: (*a*) the manic-depressives would not be expected to differ from the controls on the value Achievement scale; (*b*) the manic-depressives would be expected to score lower on the California Fascism scale, and the Traditional Family Ideology Scale; and (*c*) the manic-depressives would be expected to score higher in need Achievement phantasy and the performance tasks.

METHOD

Subjects

The subjects (*S*s) were 24 remitted manic-depressives and 30 nonpsychiatric controls. All *S*s were male white veterans, aged 25 to 65, with at least sixth-grade reading ability; most were born and reared in the Southeastern United States. Remitted manic-depressives were used in this investigation in order to study manic-depressive character undistorted by the confounding effects of severe affective disturbance and treatment regimens. These cases were carefully selected on the basis of Kallman's (1953) definition of manic-depressives as "cyclic cases which show periodicity of acute, self-limited mood swings before the fifth decade of life and no progressive or residual personality disintegration before or after such episodes" (p. 130), the absence of Lewis and Piotrowski signs (Lewis & Piotrowski, 1954), and judgments of clinical remission discussed in detail elsewhere (Becker, 1958). All of the manic-depressive patients had been at the Durham VA Hospital some time during the past five years.

TABLE 1
COMPARISON OF MANIC-DEPRESSIVES AND CONTROLS ON MATCHING VARIABLES

Measure	Manic-Depressives (N 24)			Controls (N 30)			U	p[a]
	Median	Mean	SD	Median	Mean	SD		
Age	41.50	44.13	10.56	38.50	43.17	12.10	344	.78
Education	12.60	12.13	2.91	11.50	12.17	3.66	343.5	.77
Ohio Literacy	36.50	36.50	5.00	37.67	36.50	5.98	382.5	.70

[a] A probability value of $p > .05$ (two-tailed tests) was set as the level of significance required for rejection of the null hypothesis in all of the statistical procedures of this study.

The control group consisted of 30 *Ss* who were free of psychosomatic symptoms, chronic physical disorder, and psychiatric illness. The control *Ss* were comparable to the manic-depressives for age, education, and reading ability. The medians, means, and standard deviations of the experimental and control groups on the equated variables are presented in Table 1.

Since the data for age, education, and literacy were either markedly skewed or rectangularly distributed, the Mann-Whitney *U* Test was used to determine whether the groups had been drawn from the same population with regard to the control variables. There were no significant differences between manic-depressives and controls on any of these variables.

Procedure

Nine experimental tasks were administered to each *S* in an individual testing session which required approximately 90 min. The following orienting instructions were given in a conversational manner:

> The purpose of the tasks you're about to take is to allow evaluation of some important attitudes and behaviors that seem to have a lot to do with how a person gets along outside the hospital. We want to compare people who have had some emotional difficulties, but recovered or improved, with others who have never been hospitalized for this reason.

The intent of these instructions was to stimulate interest in the tasks without unduly arousing an achievement or affiliative orientation.

The experimental procedures were introduced by asking the *Ss* to read the instructions for the Hildreth Mood Scale (Hildreth, 1946) and to fill it out. They were then administered a four-card need Achievement measure ("Inventors," "Boy in Checkered Shirt," "Trapeze Artists," and "Architects") which deviated from the standard procedure in that *Ss* were tested individually rather than in a group and the usual 5-min. time limitation per story was not imposed. Most *Ss*, however, finished the individual stories within a 5-min. period.

After completing the need Achievement measure, the *Ss* were given the Ohio Literacy Test (Foster & Goddard, 1924).

Next, the three attitude scales—value Achievement scale (DeCharms et al., 1955), F scale (Adorno, Frenkel-Brunswik, Levinson, & Sanford, 1950), and TFI Scale—(Levinson & Huffman, 1955), were administered together. Since instructions for each of these scales are similar, only one set of instructions was given. After the attitude scales were completed, the IPIT was given.[4]

The two performance measures were the final tasks administered. The Xs-in-Circles task (Atkinson & Reitman, 1956) was given with task oriented instructions, followed by the Additions task (Williams, 1955). Achievement oriented instructions were used with the latter task, stressing speed, accuracy, and competition.

[4] Because of the difficulty in generating hypotheses within a need or value Achievement framework, data for this measure are not reported in this study.

The finding that the manic-depressives did not differ significantly in affective state and that as a group they were not more affectively variable at the time of testing than the controls supports the clinical judgment that the manic-depressive *Ss* were in remission. This finding also supports the contention that any obtained differences on the experimental measures are not primarily due to differences in momentary affective state.

The medians, means, and standard deviations of the experimental measures are reported in Table 3. The differences between the manic-depressives and the controls in the experimental measures were evaluated by *t* tests when the data were approximately normally distributed and by nonparametric Mann-Whitney *U* tests when the data departed markedly from a normal distribution. The findings relative to the hypothesized performance of manic-depressives as value Achievers were as follows:

Hypothesis 1: The scores for the manic-depressives on the value Achievement scale were significantly ($U = 214$, $p < .02$) higher than those of the controls. This finding confirms Hypothesis 1; the manic-depressives placed a strong, positive, conscious valuation on achievement.

Hypothesis 2: Manic-depressives obtained a significantly ($t = 2.05$, $p < .05$) higher mean score than controls on the F scale. This finding confirms Hypothesis 2; the manic-depressives performed in a manner similar to that of high value Achievers on the F scale. High scorers on this scale tend to be characterized by rigidly conventional authoritarian attitudes, conformity, intolerance of ambiguity, and social imperceptiveness.

Hypothesis 3: The mean score of the manic-depressives on the TFI scale was significantly ($t = 2.74$, $p < .01$) higher than was that of the controls. This confirms Hypothesis 3; the manic-depressives performed in a manner similar to that which would be expected of high value Achievers. High scorers on the TFI scale are said to be characterized by attitudes of submission to authority, extreme emphasis on discipline, and rigidly defined roles and statuses for family members.

Hypothesis 4: The distribution of scores for the manic-depressives on the need Achievement measure was not different from the distribution of the controls ($U = 368$). The fact that manic-depressives do not score as would be expected if they were high need Achievers is not inconsistent with their being high value Achievers and supports Hypothesis 4.

Hypothesis 5: The scores of the manic-depressives were not different from those of the controls on either the Xs-in-Circles Task ($U = 372$) or the Additions Task ($U = 309$). The manic-depressives did not

perform more rapidly under either the "task" or "achievement" oriented instructions. If the manic-depressives were high need Achievers, they would be expected to score higher on either one or both of these tasks. The finding that manic-depressives did not score higher is not inconsistent with their performing like high value achievers and supports Hypothesis 5.

RESULTS

The Hildreth scale served the twofold purpose of checking the reliability of the psychiatric judgment of clinical remission and affording a basis for evaluating the effect of mood state on the experimental measures. The means and variances of the mood scale scores for the manic-depressives and controls are presented in Table 2. A t test of the difference between the means was not significant, nor was an F test (df 23, 29) of the differences between the variances.

In general, the manic-depressives performed in a manner which would be consistent with their being high value Achievers on all of the tasks administered. On none of the measures did the manic-depressives perform as would be expected of high need Achievers.

Effects of Age and Educational Level

Age and educational level have been reported to affect several of the experimental measures used in this investigation. The effects of these variables were partially controlled by restricting the age range (25–65) and by requiring a sixth-grade reading ability. In addition, the effects of age and education within the range of subject variability were examined. Mean scores on each of the experimental measures for manic-depressives and for controls above their respective median ages were compared with mean scores for those below the median ages; similarly mean scores for *S*s who exceeded the median educational level were compared with scores for *S*s who were below the median educational level (Becker, 1958).

TABLE 2
COMPARISON OF MANIC-DEPRESSIVES AND CONTROLS FOR SUBJECTIVE MOOD STATE

Statistic	Manic Depressive (N 24)	Control (N 30)	p
Mean	24.69	23.82	*ns*
Variance	17.31	11.36	*ns*
Range	19.0–35.5	18.3–30.3	

TABLE 3
COMPARISON OF MANIC-DEPRESSIVES AND CONTROLS ON SIX EXPERIMENTAL MEASURES

Measures	Manic Depressives (N 24)			Controls (N 30)			p
	Median	$\bar{x}$	SD	Median	$\bar{x}$	SD	
v Achievement	4.83	4.86	1.00	4.23	4.16	.67	<.02
F scale	4.62	4.57	.87	4.17	4.09	.83	<.05
TFI scale	4.29	4.19	.73	3.76	3.67	.67	<.01
n Achievement	4.25	5.38	3.96	5.00	5.77	4.68	.88
Xs-in-Circles	562.00	571.17	122.28	560.00	573.83	114.92	.83
Additions[a]	505.00	534.58	230.17	388.50	497.13	220.	.37

[a] Scores are in total seconds to complete the 200 addition problems.

Value Achievement scores increased significantly with age for the manic-depressives (p = <.05) but not for the controls. Whereas level of education was inversely related to scores on the F scale (p = <.05), and TFI (p = <.05) for the controls, no significant relation was obtained for the manic-depressives.

In order to determine whether the controls were representative of nonpsychiatric populations generally, their performance was compared with normative data for comparable age and educational levels. No significant differences were found (Becker, 1958).

DISCUSSION

Manic-depressives in this study performed like high value Achievers with respect to achievement striving, attitudes toward authority, and conventional beliefs. In the manic-depressives as in the high value Achievers, achievement and conformity behavior seem to have the goal of gaining the approval of others rather than of satisfying internalized standards. The defensive conforming nature of the manic-depressive's achievement strivings, and the rigidity of his highly stereotyped, conventional attitudes are further supported by the effects of age and educational level on these variables. Whereas achievement values were independent of age for the control *S*s, the strong conscious valuation of achievement for manic-depressives increased with age. Pros-

pects of diminished capacity with increasing age in those excessively dependent on external sources for self-esteem may cause the manic-depressive to strengthen his achievement values since he perceives them as a means of gaining needed approval. With respect to rigidity of attitudes, the manic-depressives' high authoritarian attitudes as measured by the F scale and TFI are unmodified by education. In the general population, authoritarian attitudes decrease as a function of educational level.

The results of the present study, in addition to indicating that the manic-depressive performs as a high value Achiever, served to verify empirically findings of Cohen et al., concerning a number of manic-depressive character traits. The reformulation of psychoanalytically derived descriptions of manic-depressive character within the framework of McClelland's experimentally derived concepts of achievement striving provided for a more objective and rigorous evaluation of these traits. The generality of the findings of Cohen et al., were extended by the present results to a somewhat different subculture of predominately white Southern Protestants. Furthermore, empirical verification of a number of behavior traits attributed to manic-depressives lends inferential support to their etiological formulation.

The findings of relatively specific adult achievement and conformity traits in manic-depressives suggests that hereditary-constitutional factors, are probably not sufficient in themselves to account for manic-depressive psychosis. These findings are, to a certain extent, compatible with speculations by traditional psychoanalytic writers (Fenichel, 1945) about the etiology of manic-depressive psychosis in terms of oral character structure. The conformity behavior of manic-depressives, which obtains the approval of others for them, is consistent with the oral character's intense dependency on others. Oral character explanations of manic-depressive psychosis, however, typically neglect the specification of socio-environmental factors. The findings of the present study support those of several recent inquiries (Finley & Wilson, 1951; Gibson, 1958; Wilson, 1951) which are in accord with the stress by Cohen and her collaborators on the importance of interpersonal experience in the etiology of manic-depressive character.

The results of the present study indicated that McClelland's Achievement concepts provided a useful framework within which to study pathological character. The utility of the value Achievement concept, however, was limited by its relatively undeveloped status. Greater research leverage could be achieved by translating manic-depressive character into a personality type whose characteristics, correlates, and genesis have been more comprehensively studied. The concept of the authoritarian personality, studied by a wide range of psychological techniques (Adorno, Frenkel-Brunswik, Levinson, & Sanford 1950; Titus & Hollander, 1957), seems to offer considerable research potential in this respect. In the present study, manic-depressives scored high on measures of authoritarianism and conventional attitudes. This finding suggests that such correlates of high authoritarianism as intolerance of ambiguity, social imperceptiveness, and various types of rigidity may be productively explored with manic-depressives.

SUMMARY

The present study investigated the achievement and conformity behavior of manic-depressives within the framework of McClelland's concepts of achievement striving. On the basis of similarities between McClelland's high value Achievers and manic-depressives as described by Cohen et al., specific hypotheses about the performance of manic-depressives on tasks relevant to value Achievement were formulated. Other tasks were included to test the possibility that the achievement strivings of manic-depressives may be similar to those of the high need Achiever. Twenty-four carefully selected manic-depressives were matched with thirty nonpsychiatric controls for age, education, and literacy level. All of the subjects were male white veterans between the ages of 25 and 65, with a minimum of sixth-grade reading ability; most were born and reared in the South. Nine experimental measures were used.

The manic-depressives scored significantly higher than the controls on value Achievement, the F scale, and the TFI scale. There were no differences in the scores of manic-depressives

and controls on a need Achievement measure or on either of two performance tasks. Thus, the manic-depressives performed as high value Achievers in that they showed a strong conscious valuation of achievement and highly authoritarian and conventional attitudes. The effects of age and educational level on these measures added further support to the performance of manic-depressives as value Achievers. On none of the measures did the manic-depressives perform as high need Achievers.

REFERENCES

ADORNO, T. W., FRENKEL-BRUNSWIK, E., LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality.* New York: Harper, 1950.

ARIETI, S. The decline of manic-depressive psychosis: Its significance in the light of dynamic and social psychiatry. *The scientific papers of the 113th annual meeting of the Amer. Psychiat. Assn. in summary form,* No. 119, Wash., D. C.: Amer. Psychiat. Assn., 1957.

ATKINSON, J. W., & REITMAN, W. R. Performance as a function of motive strength and expectancy of goal attainment. *J. abnorm. soc. Psychol.,* 1956, **53,** 361–366.

BECKER, J. An analysis of manic-depressive character in terms of achievement and conformity. Unpublished doctoral dissertation, Duke Univer., 1958.

BROWN, R. W. A determination of the relationship between rigidity and authoritarianism. *J. abnorm. soc. Psychol.,* 1953, **48,** 469–476.

COHEN, MABLE B., BAKER, GRACE, COHEN, R. A., FROMM-REICHMANN, FRIEDA, & WEIGART, EDITH V. An intensive study of twelve cases of manic-depressive psychosis. *Psychiatry,* 1954, **17,** 103–137.

DECHARMS, R., MORRISON, W. H., REITMAN, W., & MCCLELLAND, D. C. Behavioral correlates of directly and indirectly measured achievement motivation. In D. C. McClelland (Ed.), *Studies in motivation.* New York: Appleton-Century-Crofts, 1955. Pp. 414–423.

FENICHEL, O. *The psychoanalytical theory of neuroses.* New York: Norton, 1945.

FINLEY, CECILE B., & WILSON, D. C. The relation of the family to manic-depressive psychosis. *Dis. nerv. System,* 1951, **12,** 39–43.

FOSTER, V. H., & GODDARD, H. H. The Ohio-Literacy Test. *Pedagogical Seminary,* 1924, **31,** 340–351.

GIBSON, R. W. The family background and early life experience of the manic-depressive patient. *Psychiatry,* 1958, **21,** 71–90.

HILDRETH, H. M. A battery of feeling and attitude scales for clinical use. *J. Clin. Psychol.,* 1946, **2,** 214–221.

KALLMAN, F. J. *Heredity in health and mental disorder.* New York: Norton, 1953.

LEVINSON, D. J., & HUFFMAN, PHYLLIS E. Traditional family ideology and its relation to personality. *J. Pers.,* 1955, **23,** 251–273.

LEWIS, N. D. C., & PIOTROWSKI, Z. A. Clinical diagnosis of manic-depressive psychosis. In P. H. Hoch & J. Zubin (Ed.), *Depression.* New York: Grune & Stratton, 1954.

MCCLELLAND, D. C. Some social consequences of achievement motivation. In M. R. Jones (Ed.), *Nebraska symposium on motivation, 1955.* Lincoln: Univer. Nebraska Press, 1955. Pp. 41–65.

MCCLELLAND, D. C., ATKINSON, J. W., CLARK, R. A., & Lowell, E. L. *The achievement motive.* New York: Appleton-Century-Crofts, 1953.

RIESMAN, D., GLAZER, V., & DENNEY, R. *The lonely crowd.* New Haven: Yale Univer. Press, 1950.

TITUS, H. E., & HOLLANDER, E. P. The California F Scale in psychological research: 1950–1955. *Psychol. Bull.,* 1957, **54,** 47–65.

WENDT, H. W. Motivation, effort, and performance. In D. C. McClelland (Ed.), *Studies in motivation.* New York: Appleton-Century-Crofts, 1955. Pp. 448–459.

WILLIAMS, J. E. Mode of failure, interference tendencies, and achievement imagery. *J. abnorm. soc. Psychol.,* 1955, **51,** 573–580.

WILSON, D. C. Families of manic-depressives. *Dis. nerv. System,* 1951, **12,** 362–369.

WINTERBOTTOM, M. R. The relation of childhood training in independence to achievement motivation. Unpublished doctoral dissertation, Univer. of Michigan, 1953.

(Received February 16, 1959)

Journal of Abnormal and Social Psychology
Vol. 60, No. 3, 1960

THE INTERPRETATION OF OPINION STATEMENTS AS A FUNCTION OF RECIPIENT ATTITUDE[1]

MELVIN MANIS

Veterans Administration Hospital, Ann Arbor, Michigan

ALTHOUGH there is general agreement that a person's attitudes influence his interpretation of opinion statements, the direction and magnitude of this influence is as yet unclear. One theory, which has received perhaps its clearest statement from Festinger (1957), leads to the prediction that in interpreting statements of opinion, the recipient minimizes the apparent disparity between his own views and the views of the communicator. Presumably, so doing reduces *dissonance* and the tension that it commonly produces. Message distortion of this type has been reported by a number of investigators (Allport & Postman, 1952; Cooper & Jahoda, 1947; Edwards, 1941; Prothro, 1955).

The views of Hovland, Harvey, and Sherif (1957) and their associates stand in marked contrast to Festinger's (1957) position, although to the best of the writer's knowledge there has been scant recognition of this fact and a critical comparison of the two theories has not been attempted. Instead of focusing attention on the motivational consequences of communicator-recipient disagreement, the Hovland-Sherif group rely heavily on the results obtained in psycho-physical experiments. They cite a number of studies which show that in many judgmental tasks, introducing anchors (standards) which are at the end of the series of stimuli to be judged, or slightly removed from the end, results in a displacement of the scale of judgment *toward* the anchor (the so-called "assimilation effect"). When the reference point is too remote, however, the displacement is *away* from the anchor (the "contrast effect"). Extrapolating from these findings, Hovland, Harvey, and Sherif (1957) are led to the possibility that "when *S*s have established attitudes and are personally involved in a controversial social issue, their 'own stand' functions as the major anchorage affecting reaction to and evaluation of communication. In this case, communication near *S*'s own stand would be assimilated to it, while communication at variance with *S*'s own stand would be displaced still further away (contrast effect). Whether assimilation or contrast effects appear would be a function of the relative distance between *S*'s own stand and the position of communication." These investigators further speculate that contrast effects should begin to appear when the message advocates a position that is unacceptable to the recipient; if the message advocates a position that is within *S*'s "latitude of acceptance," or slightly beyond this range, assimilation effects are predicted.

Prothro's results (1957) are in general accord with these principles, as are the findings of Hovland, Harvey, and Sherif (1957), who originally proposed the theory; in the latter study, however, the data obtained from *S*s who were expected to show the assimilation effect offered only weak support for the theory. Other studies (Hovland & Sherif, 1952; Sherif & Hovland, 1953) have reported contrast effects, without revealing any evidence of assimilation under conditions of minimal conflict between the views of communicator and recipient.

In addition to their work relating attitudes to message interpretation, Sherif and Hovland (1953) have shown that attitudes can affect the extension or constriction of the scale that the individual uses in judging opinion statements. These investigators asked white and Negro *S*s, who were chosen because of their conflicting social attitudes, to sort statements about the Negro into as many categories as they considered necessary; Negro *S*s used fewer categories than did the whites. The re-

[1] At the time that this study was conducted, the author held a Public Health Service career-teacher grant at the University of Pittsburgh. The author would like to thank E. Lowell Kelly and Milton Rokeach for their editorial suggestions in the preparation of this manuscript; the efforts of Gloria T. Horn and David Casey, who helped in the analysis of the data, are also gratefully acknowledged.

All statements are those of the author and do not necessarily represent the opinions or policy of the Veterans Administration.

sults were interpreted as indicating that "intensely involved individuals with an extreme position are more likely to adopt a constricted scale (with few categories)."

The present experiment represents an attempt to obtain additional data pertaining to the effect of recipient attitude on (*a*) message interpretation and (*b*) scale constriction.

METHOD

Messages

A class of 35 undergraduate students wrote anonymous statements explaining their views on college fraternities. From this pool, four profraternity, four antifraternity, and four relatively neutral messages were selected for inclusion in the experiment. The primary criterion in selecting the messages was the authors' ratings of the concept *college fraternities* on the following evaluative scales: kind–cruel, valuable–worthless, honest–dishonest, fair–unfair, nice–awful, and good–bad. The six scales were presented graphically in a seven-interval format and were arbitrarily rotated to reduce response set, i.e., the "good" pole was on the left for half the scales and on the right for the other half. The obtained ratings were summed across scales, after a value of 1 had been assigned to the "good" end and 7 to the "bad" end of each scale. Previous research by Osgood and his collaborators (Osgood, Suci, & Tannenbaum, 1957) has shown that attitude scales of this type have satisfactory reliability and correlate highly with the more conventional Thurstone and Guttman scales. The authors' ratings were obtained immediately after their messages had been completed.

Recipients

The experimental messages were presented to 70 undergraduate psychology students. Before reading the messages, the recipients' attitudes toward fraternities were assessed by means of the rating scale procedure. They then read the experimental passages and attempted to predict each writer's ratings of college fraternities, again using the six evaluative rating scales; these predictions were summed across scales to yield a measure of each communicator's judged position on the attitude continuum. The passages were assembled in a constant, chance-determined order, with the restriction that no two messages expressing a similar viewpoint could appear in succession.

In order to compare the judgments of *S*s who held varying attitudes, the total sample of recipients was divided into quintiles on the basis of their attitude scores; there were 14 *S*s in each quintile.

RESULTS AND DISCUSSION

Figure 1 presents the mean responses of the five recipient-groups to the profraternity, neutral, and antifraternity messages; these responses represent the recipients' predictions as to the views held by the writers. The results for each type of message are shown separately, with the recipient-groups arranged from left to right, in order of decreasing favorability toward fraternities; predictions that are high on the ordinate reflect a belief that the communicator *disliked* fraternities. An analysis of variance revealed that in responding to the antifraternity passages, the recipient groups differed significantly ($p < .05$); for the profraternity and the neutral messages, the group differences fell short of significance ($p < .10$).[2]

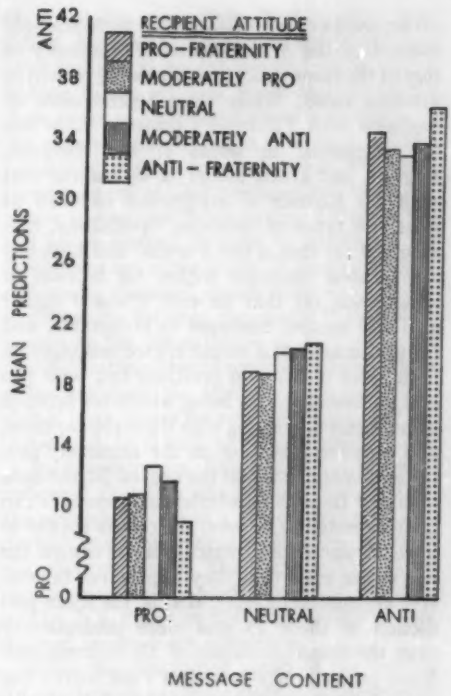

FIG. 1. Mean predictions as a function of attitude.

Interestingly enough, the responses to the "pro" and the "anti" messages show curvilinear trends that are essentially mirror images

[2] With 4 and 65 degrees of freedom, an F of 2.51 was needed for significance at the .05 level. For the antifraternity messages, F equalled 2.79; for the profraternity messages, F equalled 2.49; for the neutral messages, F equalled 2.29.

Application of the analysis of variance for trends (Grant, 1956; Grandage, 1958) revealed significant quadratic components ($p < .01$) for the profraternity and antifraternity messages and a significant linear component ($p < .01$) for the neutral messages.

of one another. In both cases, the more strongly committed the *S*, the greater his tendency to regard the communicators as holding relatively extreme views. While these results seem at variance with Festinger's position, they can be interpreted in terms of the Hovland, Harvey, and Sherif model, if we assume that each *S*'s latitude of acceptance included at least two types of messages. Specifically, it is assumed (*a*) that a pro *S* would find both pro and neutral messages within his latitude of acceptance, (*b*) that an anti *S* would regard anti and neutral messages as acceptable, and (*c*) that a neutral *S* would regard messages on both sides of his own position, i.e., both pro and antimessages, as being within his latitude of acceptance. Starting with these assumptions, and focusing attention on the extremely pro, the extremely anti, and the neutral *S*s, the data obtained from the profraternity messages can be interpreted as follows: the pro *S*s appear to have displaced the communicators *toward* the end of the scale that they themselves favored (the assimilation effect); that is, the mean prediction of these *S*s was more profraternity than the mean prediction of all *S*s combined. Since profraternity messages were within the pro's latitude of acceptance, a shift of this sort would be consistent with the Hovland, Harvey, and Sherif theory. The anti *S*s seem to have displaced the procommunicators *away* from their own anti stands (the contrast effect),[3] a result which is also consistent with the theory, since profraternity messages would doubtless be outside the anti's latitude of acceptance. Finally, the neutral *S*s seem to have shifted the procommunicators toward the midpoint of the scale (assimilation); this too is in accord with the theory, if we assume (as above) that neutral *S*s would regard pro messages as acceptable. Similar reasoning can be applied to the data obtained from the antifraternity messages.

The predictions made by the moderately pro- and moderately antifraternity *S*s, in response to the two extreme groups of messages, are somewhat more difficult to interpret. These *S*s apparently responded much as the neutrals did, in that their general tendency was to assimilate both pro- and antifraternity messages toward their own positions. In order to integrate these data with the Hovland-Harvey-Sherif theory, one would have to assume that both extreme viewpoints were within the "moderates'" latitude of acceptance, or slightly beyond this range.

On the neutral messages, there was a linear relationship between attitude and interpretation of messages; *S*s who favored fraternities felt that the neutral communicators were more profraternity than did *S*s who opposed them. While this assimilation trend is clearly in accord with Festinger's viewpoint, it is also consistent with the Hovland, Harvey, and Sherif model, if we assume (as above) that neutrality was an acceptable position to all recipients.

It is interesting to contrast the results of the present experiment with those reported by Segall (1959). This investigator had university students make absolute judgments of statements about fraternities and found virtually no differences between judges of varying attitude. He reasons that his negative results may have occurred because of the relatively narrow range of opinions represented in his *S*s, and this may indeed be a crucial difference between the two studies, for the present data were collected on a big city campus, where the fraternity system was probably more controversial than it was in the setting that Segall describes. It is also possible that the different methods employed in the two studies were crucial in determining their conflicting results; that is, while Segall's *S*s were asked to judge the extremity of opinion statements, *S*s in the present experiment did not judge the statements; instead, they predicted the positions advocated by the persons who had made the statements. It may be that this latter set is more likely to call the *S*'s own views into play as an implicit standard and determinant of his response.

Attitude and Scale Constriction

Scale constriction was measured by computing a standard deviation for each *S*'s predictions over the set of 12 messages. Figure 2 presents the mean standard deviation for each of the five groups of recipients. The groups are again arranged in order of decreasing favor-

[3] This statement and subsequent statements regarding judgmental shifts are based on a comparison of the mean prediction given by a specific group of *S*s (here, the antis), and the mean prediction of all *S*s combined.

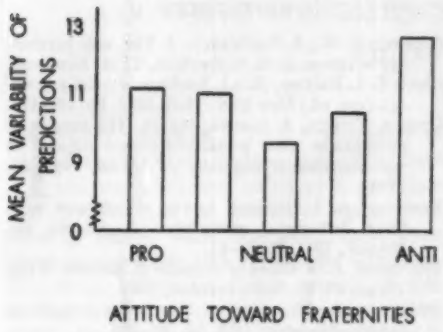

FIG. 2. Variability of predictions as a function of attitude.

ability toward fraternities, with the pro-fraternity *Ss* on the left and the antifraternity *Ss* on the right. An analysis of variance revealed that the difference between the group means was significant at the .025 level. As will be seen, the relationship between attitude and scale constriction is curvilinear. *Ss* with relatively neutral views perceived the messages as reflecting a narrower range of opinion than did *Ss* who were committed on the fraternity issue; apparently the neutral *Ss* felt that there were only minor differences between the various communicators. This trend is perhaps seen most clearly if Fig. 2 is "folded over," resulting in the formation of three recipient-groups: the highly committed *Ss*, the moderate *Ss*, and the neutral *Ss*. The mean standard deviations for these groups are 11.9, 10.7, and 9.6 respectively.

It is likely that these constriction differences represent a "byproduct" of the displacement trends that were discussed in connection with Fig. 1. That is, the neutral and moderate *Ss* may show a more constricted scale of judgment than the committed *Ss* because they tended to displace the communicators *toward* the midpoint of the scale, while the committed *Ss* displaced the communicators *away* from the midpoint.

Although the results obtained from the pro and the anti messages (see Fig. 1), and the results shown in Fig. 2, could not readily be attributed to random error, they could have been due to differences in response set; more specifically, the relatively uncommitted *Ss* may have been unwilling to use the extremes of the scales, regardless of what they were rating. A response set of this sort could have caused these *Ss* to (*a*) respond in a neutral fashion to the attitude scales and (*b*) reduce their range of judgments when responding to the experimental messages. To check on this possibility, the five recipient groups were compared in terms of their ratings of two unrelated concepts—jazz and labor unions. These ratings had been obtained in order to familiarize the *Ss* with the use of the scales. If the constriction differences discussed above were, in fact, due to differences between the groups in their willingness to use the extremes of the scales, the neutrals' ratings should have shown only slight deviations from the midpoint, the moderates should have shown greater deviations, and the committed *Ss* should have been most willing to use the extremes. Examination of the data indicated that this was *not* the case; in fact, the neutral *Ss* showed a greater willingness to deviate from the midpoint than did either of the other two groups. The relationship between attitude and scale constriction, which is shown in Fig. 2, is therefore not attributable to generalized response set differences.

While the constricted range of the neutral *Ss*' predictions was quite marked in the present study, the obtained results seem somewhat at variance with the interpretation that Sherif and Hovland (1953) offer for their data. As noted above, these investigators found that Negro *Ss* used a relatively constricted scale in judging items dealing with the Negro, and concluded that they did so because of their intense involvement with the experimental issue. The results of the present experiment seem to contradict this interpretation, if we make the assumption that the neutral *Ss* were probably *less* intensely involved in the fraternity issue than the committed *Ss*. While these discrepant findings may be attributable to differences in experimental procedure or topic, it may be that the Sherif and Hovland results were not due to their *Ss*' varying degrees of involvement in the experimental task; instead, they may reflect the operation of cultural, educational, or intellectual factors which might have caused the Negro *Ss* to use a relatively constricted scale, regardless of the stimulus materials that they were asked to judge.

SUMMARY

Five groups of college students, who varied in their attitude toward fraternities, were presented with the written statements of 12 communicators, who had been asked to explain their views on fraternities. There were three types of communicators: those who favored fraternities, those who opposed them, and those who were relatively neutral. After reading each message, the recipients were asked to indicate the communicator's attitude toward fraternities, as defined by his response to an attitude scale.

In responding to the profraternity and the antifraternity messages, there was a curvilinear relationship between attitude and judgment: in both cases, the committed *Ss* perceived the communicators as occupying more extreme positions than did the neutral *Ss*. In responding to the neutral messages, the relationship between attitude and judgment was linear: *Ss* who favored fraternities felt that the neutral communicators were more profraternity than did the *Ss* who opposed them. These results were interpreted as being generally consonant with a theory developed by Hovland, Harvey, and Sherif (1957).

A related finding was the fact that the neutral *Ss* used a relatively constricted scale of judgment; they clustered their predictions within narrower limits than did the committed *Ss*, apparently feeling that the messages did not reflect a particularly wide range of opinion. If we make the assumption that the neutral *Ss* were probably *less* intensely involved in the fraternity issue than the committed *Ss*, these results appear to contradict Sherif and Hovland's contention that intensely involved individuals with an extreme position are likely to adopt a constricted scale of judgment.

REFERENCES

ALLPORT, G. W., & POSTMAN, L. J. The basic psychology of rumor. In G. E. Swanson, T. M. Newcomb, & E. L. Hartley, (Eds.), *Readings in social psychology.* (rev. ed.) New York: Holt, 1952. Pp. 160–170.

COOPER, EUNICE, & JAHODA, MARIE. The evasion of propaganda: How prejudiced people respond to anti-prejudice propaganda. *J. Psychol.* 1947, **23**, 15–25.

EDWARDS, A. L. Political frames of reference as a factor influencing recognition. *J. abnorm. soc. Psychol.*, 1941, **36**, 34–61.

FESTINGER, L. *A theory of cognitive dissonance.* White Plains, N. Y.: Row, Peterson, 1957.

GRANDAGE, A. Orthogonal coefficients for unequal intervals. *Biometrics*, 1958, **14**, 287–289.

GRANT, D. A. Analysis-of-variance tests in the analysis and comparison of curves. *Psych. Bull.*, 1956, **53**, 141–154.

HOVLAND, C. I., & SHERIF, M. Judgmental phenomena and scales of attitude measurement: Item displacement in Thurstone scales. *J. abnorm. soc. Psychol.*, 1952, **47**, 822–832.

HOVLAND, C. I., HARVEY, O. J., & SHERIF, M. Assimilation and contrast effects in reaction to communication and attitude change. *J. abnorm. soc. Psychol.*, 1957, **55**, 244–252.

OSGOOD, C. E., SUCI, G. J., & TANNENBAUM, P. H. *The measurement of meaning.* Urbana: Univer. Illinois Press, 1957.

PROTHRO, E. T. The effect of strong negative attitudes on the placement of items in a Thurstone scale. *J. soc. Psychol.*, 1955, **41**, 11–17.

PROTHRO, E. T. Personal involvement and item displacement on Thurstone scales. *J. soc. Psychol.*, 1957, **45**, 191–196.

SEGALL, M. H. The effect of attitude and experience on judgments of controversial statements. *J. abnorm. soc. Psychol.*, 1959, **58**, 61–68.

SHERIF, M., & HOVLAND, C. I. Judgmental phenomena and scales of attitude measurement: Placement of items with individual choice of number of categories. *J. abnorm. soc. Psychol.*, 1953, **48**, 135–141.

SHERIF, M., TAUB, D., & HOVLAND, C. I. Assimilation and contrast effects of anchoring stimuli on judgments. *J. exp. Psychol.*, 1958, **55**, 150–155.

(Received February 27, 1959)

Journal of Abnormal and Social Psychology
Vol. 60, No. 3, 1960

COGNITIVE CONSISTENCY AND ATTITUDE CHANGE

WILLIAM J. McGUIRE
University of Illinois

THERE has been a recent resurgence of interest among psychological researchers in the predictive value of a postulate that people tend to maintain logical consistency among their cognitions (and even between cognitions and more gross behavior). The most influential single source of this revived interest is probably Heider's (1946, 1958) "balance" theory, while Festinger's (1957) theory of cognitive dissonance is probably the most stimulating elaboration to date. Most of the approaches that employ this postulate recognize, implicitly or explicitly, that the tendency towards consistency is not absolute. Its expression is limited by other, potentially conflicting forces in the person and by the inadequacies of his cognitive apparatus, even in the absence of conflicting forces, to achieve full consistency. These coordinate hypotheses of cognitive consistency and its predictable limitations have been applied successfully to the prediction of behavior in many areas. The work reported here involves two applications in the area of persuasion.

The first of these applications involves what will be called the "Socratic method" of persuasion. The persuasion technique usually studied in the social sciences involves inducing a change in a person's opinion by presenting him with a persuasive message from an external source, either through a "mass" one-way communication channel or in a face-to-face reciprocal communication network. The postulate of cognitive consistency suggests that persuasion could be effected by the quite different technique of eliciting the persuasive material from the person's own cognitive repertory, rather than presenting it from outside. Specifically, this technique would involve asking the person his opinions on logically related issues, thus sensitizing him to any inconsistencies that exist among his stands on these issues and producing a change towards greater mutual consistency. The technique has an obvious affinity to the "Socratic method" used in the Platonic dialogues (Jowett, 1937).

The derivation of the foregoing hypothesis requires, besides the postulate of a tendency to maintain cognitive consistency, the assumptions that various distorting factors (of which "wishful thinking" is given primary attention in this study) have introduced inconsistency among the initially elicited set of opinions, and that the temporally contiguous elicitation of these opinions brings the tendency towards consistency more strongly to bear upon them.

Part of the study described below was designed to test the hypothesis that a person tends to change his opinions on logically related issues in the direction of greater mutual consistency when he is merely asked to state these opinions in close temporal contiguity. Several additional hypotheses regarding the focus of the Socratic effect, that is, the type of opinion most likely to be adjusted in case of inconsistency, were also tested.

The second application of the cognitive consistency postulate to the area of persuasion in the present experiment dealt with the logical repercussions of persuasive communications. The effectiveness of a persuasive message is usually gauged by the amount of change in opinion that it induces on the issue with which it explicitly deals. From the postulated consistency tendency, however, it would follow that an induced change on the explicit issue tends also to produce changes on logically related (unmentioned) issues such as would tend to maintain internal consistency among beliefs.

From the additional, limiting assumption that there is a certain amount of inertia in the person's cognitive apparatus, two additional hypotheses were derived. It was hypothesized that the consistency-maintaining change expected on remote issues is less than the amount logically required for complete consistency with the induced changes on the explicit issues. Further, it was hypothesized that inertia would also result in temporal effects, such that the impact of the message on the remote issue occurs only gradually, the opinion on the

remote issue continuing to change in the logically required direction for some time after receipt of the persuasive message.

Since most of these hypotheses involve predictions of fine graduations of change towards or away from a reference point of precise cognitive consistency, we need a quantitative definition of consistency. Such a definition is provided by a model combining formal logic and probability theory, the relevant aspects of which are summarized below.

Consider three issues so interrelated that a person taking Stands a and b on two of these issues would, if he were to be consistent, be logically required to take Stand c on the third. That is, the three stands are in a syllogistic relationship, with c following as a valid conclusion from the conjunction of a and b. If the person's opinions on these issues are obtained by having him indicate his adherence to each of the Stands, a, b, and c, on a probalistic scale, we can specify that to be completely consistent, these probabilistically scaled opinions must be interrelated in the form

$$p(c) = p(a \,\&\, b) + p(k)(1 - p(a \,\&\, b)) \qquad [1]$$

where $p(k)$ is his opinion of the probability of c on bases other than the conjunction of a and b. It can also be specified that if we induce a change of opinion on a and/or b (for example, by a persuasive communication) while $p(k)$ does not change appreciably, then the required change on c necessary for maintaining logical consistency among the opinions is

$$\Delta p(c) = \Delta p(a \,\&\, b) \qquad [2]$$

which, assuming that a and b are independent events i.e., that $p\,(a \,\&\, b) = p(a)\,p(b)$ becomes

$$\Delta p(c) = \Delta p(a)p(b) + \Delta\, p(b)p(a) + \Delta p(a)\Delta p(b) \qquad [3]$$

where $p(a)$ and $p(b)$ refer to the initial opinions, before the communication induced changes. When, in the discussion below, we refer to the "logically required change on the conclusion," we are referring to the $\Delta p(c)$ value specified in Equation 3.

METHOD

Material. The *S*s' opinions on sets of logically related issues were measured by means of a questionnaire that contained 48 statements. These statements comprised 16 sets of three syllogistically related propositions; that is, two of the propositions stated stands on a pair of issues, agreement with which would logically imply agreement with the stand taken in the proposition on the third issue.[1] The three propositions making up any one syllogism appeared in different parts of the questionnaire, being interspersed with the propositions of other syllogisms.

The *S*s were asked to indicate in probabilistic terms the extent of their adherence to the stand taken in each proposition. For this purpose each proposition was followed by a graphic scale consisting of a five-inch horizontal line marked off at half-inch intervals and calibrated 0, 10, 20. . . 90, 100. To the left of the scale was written "Very improbable end," and at the right, "Very probable end." The *S*s were asked to decide how probable they felt the truth of each statement to be, and to indicate this subjective probability by drawing a line through the scale at the appropriate point.

In a later section of the questionnaire the propositions were repeated, this time with a scale consisting of five boxes labelled from left to right "very desirable," "somewhat desirable," . . ," "very undesirable." The *S* was instructed to indicate his feeling regarding the desirability of each proposition by putting an X in the appropriate box. (As explained below, these ratings of desirability were obtained to allow identification of initial inconsistencies, due to wishful thinking, among the probability ratings.)

The questionnaires were administered during each of the three experimental sessions. Between the first and second administrations of the opinion questionnaires, a series of persuasive messages were communicated to the *S*s in mimeographed booklets.[2] There were 16 such messages, each directed at one premise of each of the 16 syllogisms. Each message argued that the proposition in question was true, i.e., was aimed at increasing the *S*'s probabalistic estimate of the statement's truth. These messages were "one-sided," 200–300 words long, employed rational appeals, and based arguments on factual evidence. They were mild in tone and attributed to a purportedly competent and dependable source.

---

[1] A complete list of all the syllogisms and persuasive messages used in the present study has been deposited with the American Documentation Institute. Order Document No. 6235, from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

One of the 16 syllogisms used was as follows. Students who violate any regulation that has been made to safeguard the lives and property of other students will be expelled; the regulations against smoking in the classrooms and corridors were made to safeguard the lives and property of the student; students who violate the regulation against smoking in the classrooms and corridors will be expelled.

[2] The text of all 16 persuasive messages can be obtained in the ADI Document No. 6235, cited in Footnote 1.

Procedure. There were three 50-min. experimental sessions. In the first, the questionnaires were administered to obtain "before communication" measures of opinions on the issues. The second session, which came one week after the first, was devoted to communicating the persuasive messages and then to a second administration of the opinion questionnaires in order to measure the immediate impact of the persuasive messages. One week later, the *S*s participated in the third session during which the questionnaires were again administered (in order to ascertain the delayed effects of the persuasive messages). At this final session the *S*s were also told the purpose of the experiment and the *E* pointed out the deceits that he had employed in the experiments and his reasons for so doing.

Until this final "catharsis," an attempt was made to conceal, or at least distract attention from, the persuasion aspect of the experiment and from the investigator's interest in cognitive consistency. The study was purported to be an investigation of people's ability to understand experts' discussions of controversial topics. The persuasive messages were alleged to be excerpts from such discussions, and after each message the *S*s were required to answer a series of comprehension questions (to emphasize the contention that we were primarily interested in studying the *S*s' understanding of the material). The opinion questionnaires were explained as required to ascertain whether the issues dealt with were actually controversial in the group being studied and whether they continued to be controversial throughout the period of the experiment.

Subjects and design. A total of 120 *S*s participated in the experiment. One-fourth of these were high school seniors and the remainder, college freshman; the ratio of men to women was about 2:1 at both grade levels. To minimize the likelihood that the consistency hypotheses would receive specious confirmation by the employment of atypically rationalistic *S*s, these *S*s were selected from a group with a rather low level of intellectual attainment. Their modal high school grade average was below the 30th percentile for all high school students in their state. Each of these students reported that he had never had a course in formal logic or in probability theory.

In order that the *S*s might have time to respond thoughtfully to the opinion items and to read the persuasive messages carefully, each *S* was required to provide data on only 8 of the 16 syllogistically related sets of propositions. On only four of these eight did he receive persuasive messages, the other four serving as no-communication "control" items. These no-communication syllogisms provided the data for testing the Socratic effect hypotheses. The syllogisms on which messages were communicated provided the data for testing the hypotheses concerning logical repercussions. Specific syllogisms were alternated between message and no-message conditions from *S* to *S*.

RESULTS AND DISCUSSION

Hypotheses Concerning the Socratic Effect

Initial inconsistencies among the opinions. In order to identify initial opinion inconsistencies we employed an auxilliary hypothesis that the subjective probabilities of the propositions would be distorted in the direction of wishful thinking. Internal evidence indicates that this assumption was a valid one. The 16 syllogisms were partitioned into two subsets: the eight in which the desirability of the conclusions was low relative to that of the premises; and the eight whose conclusions had relatively high desirabilities. Wishful thinking would tend to make the subjective probabilities of the conclusions low relative to those of their (more desirable) premises in the first subset of syllogisms, while the subjective probabilities of the conclusions in the second set should be high relative to their less desirable premises. This prediction is borne out by the subjective probability ratings obtained in the first session, as shown in Table 1. (To avoid confusing them with the significance levels, the obtained subjective probability ratings given in this report are the obtained probability ratings, multiplied by 100.) The mean subjective probability of the conclusions exceeded the mean joint probability of the premises by only 2.20 points on the 100-point scale in the first eight syllogisms, while in the other eight syllogisms (the conclusions of which had relatively high desirability) the mean subjective probability of the conclusions exceeded the mean joint probability of the premises by 17.18 points. This difference between 2.20 and 17.18 points is significant beyond the .01 level. (All significance levels reported in this paper are based on analyses of variance and two-tailed tests.) Hence, the attempted manipulation of initial inconsistency by use of the tendency toward wishful thinking was successful.

The main effect. The changes in opinion obtained in the conditions in which no persuasive message was communicated on the given issues provide the test of the main Socratic hypothesis, that when the person's opinions on related issues are elicited in close temporal contiguity, his need for consistency is brought more fully to bear on them. Hence, when these opinions are again elicited on a later occasion, they will be found to have changed towards greater mutual consistency. In view of the initial distortions from consistency produced by wishful thinking, it would follow that in the no-communication conditions in the

TABLE 1
SUBJECT PROBABILITY SCORES (MULTIPLIED BY A FACTOR OF 100) IN THE NO-COMMUNICATION CONDITIONS

	Session	Major Premises	Minor Premises	Joint of Premises	Conclusions	Excess of Conclusions over Premises
Syllogisms with conclusions more desirable than premises	1st	62.15	54.14	33.66	35.86	2.20
	Changes, 1st to 2nd	−3.04	1.45	−0.79	5.17	5.96
Syllogisms with conclusions less desirable than premises	1st	55.50	55.41	30.12	47.30	17.18
	Changes, 1st to 2nd	3.96	−1.98	2.06	0.35	−1.72

present experiment initial questioning should produce an increase in the subjective probabilities of the conclusions relative to those of the premises in the first eight syllogisms (those with the relatively undesirable conclusions); while for the second eight syllogisms, there would be a relative decline in the subjective probabilities of the conclusions.

The results bear out this prediction. As can be seen in Table 1, the gain in probability of the conclusions over that of the premises in the first eight syllogisms averaged 5.96 points. The probabilities of the conclusions of the second eight syllogisms, on the other hand, showed a loss of 1.72 points relative to the premises. This net difference of 7.68 points between means is significant at the .02 level. Hence, simply eliciting opinions on logically related issues does tend to move those opinions toward greater mutual consistency when subsequent elicitation occurs one week later.

The difference obtained between the two sets of syllogisms as regards internal changes is in the direction that would be expected on the basis of a simple regression from the initially obtained difference. In order to test whether there is a Socratic effect over and above simple regression, an additional covariance analysis was carried out. The obtained difference in change of probability scores between the two sets of syllogisms was adjusted for the difference in initial levels by means of the obtained within-set correlation between the initial probability scores and the changes therein. Even when the obtained changes are adjusted for initial levels by this analysis of covariance, the difference in changes remains significant at the .02 level, indicative that a significant Socratic effect remains, even when the effect of regression is partialed out.

The Socratic effect tends to be produced primarily by the first contiguous elicitation of the opinions, with only slight evidence of an accumulative effect from subsequent, additional elicitations. As discussed above, there was a net Socratic effect, from the first to second elicitation of the opinions, of 7.68 points. The corresponding net effect, from second to third elicitations was again in the predicted direction but amounted to only 2.62 points, which falls far short of the conventional levels of statistical significance.

Focus of the Socratic effect. An attempt was made to answer the further question of where this Socratic effect is focused; that is, which of the person's initially inconsistent opinions are most likely to be changed in bringing about the greater consistency? The present data were analyzed to ascertain whether the amount of change was related to any of three variables: the logical status of the propositions, their initial valence, or their ordinal position on the questionnaire in the first session.

As regards logical status of the proposition, we tested whether the major premise, minor, and conclusion contributed differentially to the change towards consistency. All of the syllogisms used in this study were of the "Barbara" type. Hence the major premise contained the M-P terms; the minor, the S-M terms; and the conclusion, the S-P terms (where S refers to the subject of the conclusion; P to its predicate; and M to the middle term).

Logical status of the proposition did have a significant overall impact on amount of opinion change. As can be seen in Table 1, the major premise contributed most of the change towards increased consistency, the mean change being 3.50 points on the 100-point scale. The conclusion showed the next greatest change

towards consistency, a mean of 2.41 points; and the minor premise contributed least, in fact their mean change was slightly (1.71 points) away from consistency. These differences among the three means are significant at the .05 level by an overall analysis of variance.

The overall significance of the variance among the three means is due primarily to the greater change towards consistency (averaging 2.95 points) of the propositions with the P terms (namely, the major premises and the conclusions) than of those without these P terms (namely the minor premises), which averaged a 1.71 point change away from consistency. The difference between the two classes of propositions is significant at the .03 level. It appears to be explicable on the basis of initial distortion. In these "Barbara" syllogisms, the P terms tend to have the greatest emotional valence. Therefore, by the auxiliary hypothesis of wishful thinking, the propositions that include these emotionally-tinged terms would have been the most distorted originally. The minor premises, without the P terms, are typically statements of class inclusion ("Socrates is a man") and therefore less likely to be distorted. Hence, the greater change towards consistency of the other propositions appear to be a consequence of the fact that the others had contributed more of the original inconsistency.

Two other propositional characteristics were investigated for possible relationship to the amount of change in opinion produced by the Socratic method: initial valence and initial order of presentation. It was found that somewhat more change towards consistency occurred on the highly valenced propositions—the 12 most and the 12 least desirable propositions—than on the 24 moderately valenced ones. The direction of this outcome tends to reverse the "polarity" prediction (Osgood & Tannenbaum, 1955; Tannenbaum, 1956), but it does not approach the conventional level of statistical significance. Regarding initial order, it was found that somewhat more change toward consistency occurred on those propositions which, in the first session, appeared earlier on the questionnaire than on those that appeared towards the end, but that this difference was not statistically significant. The difference has been predicted on the assumption that the later appearing propositions had already, on the first administration of the questionnaire, been brought into closer agreement with the opinions given on the propositions that had come earlier, so that less change remained to be made.

A number of other hypotheses, not tested in this study, regarding propositional characteristics that affect amount of change in cases of mutual inconsistency could be tested. Opinions on more familiar issues may be more stable than those on less familiar. Opinions on issues that show greater cognitive complexity and greater articulation with the rest of the person's cognitive system (as defined, for example, by Zajonc's [1954] procedures) might participate less in the change towards consistency than more isolated and undifferentiated opinions. Tannenbaum's (1956) finding (that when differently valenced sources and stands on issues are paired, the person's attitude on the issue shows less change than his attitude towards the source) is suggestive in this regard. Following Asch's (1952) thinking that persuasion involves a change of the issue being judged rather than change of judgement on the given issue, we might expect that in case of conflict, the greatest change would occur on the issues most easily redefined. Another hypothesis is that the relative proneness to readjustment of conflicting opinions is related to the bases for accepting the opinions. For example, following Kelman's (1956) terminology, we might conjecture that opinions based on internalization might be more prone to change, in case of internal conflict, than those based on identification or on compliance. These latter need not be consistent with one's other beliefs, because the adherence to them is based on extrinsic considerations.

Hypothesis Concerning Logical Repercussions

The results just reported regarding the Socratic effect hypotheses are all based on data in the no-communications conditions. The following results, which deal with the logical repercussions of persuasive messages are based on data from the conditions in which persuasive communications were received. The changes reported as due to the persuasive messages are computed from a baseline provided by the change from first to second session in the no-communication

condition. Since the results concerning the Socratic effect indicate certain systematic changes in opinions occur from first to second session even in the absence of persuasive messages, the message-induced changes must be computed from a control baseline that will eliminate those other changes and also any regression trends.

Immediate postcommunication changes. The induced changes of opinion on the explicit target issues (the premises) were considerable. Each message argued that the stand taken in one of the premises of each syllogism was true and that, therefore, the receiver should increase his subjective probability regarding the truth of that proposition. As can be seen in Table 2, the mean induced increase on the 16 explicit propositions immediately after the communication was 17.56 points on the 100-point scale ($p < .001$).

In order to maintain cognitive consistency, the *S*s should therefore have increased their subjective probabilities regarding the truth of the conclusions. Even though neither conclusions nor the additional premises needed for their derivation had been adverted to in the communications, a considerable immediate postcommunication increase was indeed produced on these conclusions. As can be seen in Table 2, there was an increase (as required for consistency) immediately after the messages on 14 of the 16 unmentioned conclusions, the mean increase being 5.96 points. A mean change of this magnitude is significant at the .01 level. Hence, the results do indicate a consistency tendency sufficiently strong to cause a persuasive message to have an impact on logically related issues not explicitly mentioned in the communication.

These obtained changes on the derivative issues are, however, less than the amount required for complete logical consistency as determined by substituting the amount of the obtained change on the explicit premise in Equation 3. The mean required increase on the conclusions so computed was 11.57 points, while the actually obtained increase was only 5.96 (see Table 2), a difference significant beyond the .05 level. Hence, persuasive messages, while they do produce changes on remote, unmentioned issues, produce less than the amount logically required, a finding that supports the hypothesis of a certain amount of cognitive inertia.

Delayed postcommunication changes. A considerable amount of the immediately induced persuasion on the explicit target issues had dissipated by the end of the week following the communication. As can be seen in Table 2, almost half of the immediately induced change on these explicit issues had dissipated one week later. The mean loss of 8.43 points was significant at the .02 level. On the other hand, the 9.13 points of change remaining after one week was also significant (at the .01 level), indicating that there is still a significant amount of change of opinion on the explicit issue one week after the persuasive communication.

When this loss on the premises is substituted in Equation 3 the amount of the originally induced change on the unmentioned derived issues that should logically have dissipated during the succeeding week was computed as 6.69 points (see Table 2.) The obtained loss during the interim, however, was only a negligible 1.48 points, a gain of 5.20 points relative to the explicit premises for the week ($p = .10$). In fact, by the week after the persuasive communication, the net mean change on the conclusions from their precommunication level was an increase of 4.47 points, which was almost as much as the 4.89 points logically required in view of the net changes on the premises. The trivial extent to which the change on conclusions fell short of the required amount one week after, as compared with the significant extent to which it fell short immediately after the communication, gives some support to the hypothesis that the persuasive messages had a delayed-action effect on the derived issues. The net change on the derived issues did not actually increase with the passage of time, but it fell off less sharply than would have been expected on the basis of the usual rate of temporal decay of the immediately transmitted opinion change, suggesting that continued seepage of the impact on the explicit issue almost completely overcame the opposing decay effect on the derived, unmentioned issue.

Comparison with previous studies. One methodological artifact in the present study, the use of a before-after design, may have

augmented the amount of the logical repercussions of the persuasive messages. In order to obtain each *S*'s precommunication opinions he was required to fill out, one week prior to receipt of the persuasive messages, an opinion questionnaire containing all the issues, including the auxilliary premises and conclusions not mentioned in the message itself. It is possible that having stated his opinions on these related issues during the previous week may have somewhat increased their saliency, so that they were more affected by the message than would have been the case had they not been raised the previous week.

On the other hand, the finding of logical repercussions in the present experiment is particularly convincing in view of the fact that the method included a number of conditions that have been shown in previous studies to militate against logical repercussions. In the first place, not only was the conclusion itself unmentioned in the messages but also one of the two premises necessary for its derivation was left unmentioned. Also, the *S*s employed were of a type, being below average in intelligence, that has been found (Cooper & Dinerman, 1951) to be relatively resistant to nonexplicit messages. Furthermore, no mention was made during the experiment of our interest in consistency, and the logically related propositions were widely interspersed with other propositions within the questionnaire. Finally the *S*s gave their opinions on a scale with whose computational axioms they were unfamiliar, so that they could hardly have determined the extent of their inconsistencies even had they been interested in this matter.

Hence, the negative indications reported in some previous studies for the hypothesis that persuasive communications have logical repercussions require comment. While more persuasion has typically been found on a derived issue if that issue is explicitly mentioned rather than only implied, the data reported typically do show some change on the nonexplicit issue (Hovland, Lumsdaine, & Sheffield, 1949; Hovland & Mandell, 1952). Also, the relative lack of effect on the derived issue is most evident in the immediate postcommunication periods. Hovland, Lumsdaine, and Sheffield (1949, pp. 70–71) found that the "Why We Fight" films did have an impact on more general attitudes after the passage of a greater amount of time, and have postulated that there may be need for a "sinking in" period before these derivative effects are felt. Similarly, Katz, Stotland, Patchen, and Jochen (1957) report that giving *S*s case history material regarding the roots of prejudice causes little immediate effect on their own prejudice, but after the passage of several weeks, the extrapolation of the intellectual knowledge to their opinions does occur. Furthermore, the magnitude of the logically required change on the derived issue is, as shown by Equation 3, less than that on the explicit issue. This logically consistent attenuation of the persuasive impact on derived issues due to a compounding of contingencies becomes more severe as the derived issue becomes more tenuously related to the explicit issue. For example, the more premises beside the explicit one needed for its derivation, the less the logically required change on the derived conclusion. Where the derived issues are as remote from those explicitly dealt with as was the case, for example, in the Hovland, Lumsdaine, and Sheffield "Why We Fight" studies (1949), very little change on the derived issues would be logically required even in a perfectly consistent *S*.

However, while only a small change in opinion is logically required on the derived issue, the amount actually occurring is even less. People's opinions seem to follow a "loose-link" model such that a considerable change

TABLE 2

CHANGES IN SUBJECTIVE PROBABILITIES ON EXPLICIT ISSUE AND ON DERIVED, UNMENTIONED ISSUES IMMEDIATELY AFTER AND ONE WEEK AFTER PERSUASIVE MESSAGES

	Obtained Change on Premises	Logically Required Change on Conclusions	Obtained Change on Conclusions	Discrepancies of Obtained from Required Change on Conclusions
Changes from before to immediately after	17.56	11.57	5.96	−5.61
Changes from immediately to one week after	−8.43	−6.69	−1.48	5.20

can be induced on one cognition before any pull is exerted on related cognitions. This effect, however, is at least partially corrected by the continued seepage of change from the explicit issue to the derived one as time passes after the receipt of the communication.

The question of conscious awareness arises in connection with the psychodynamics of the present effect: to what extent is it necessary that the person be consciously aware of the inconsistencies among his beliefs in order that the effects occur? In the present study several procedures were employed in an attempt to minimize the degree of conscious awareness. (For example, no mention was made to the *S*s that their consistency was being assessed; the related propositions were widely separated from one another in the questionnaire; a quantitative expression of opinion was required and the assessment of inconsistency involved a mathematical model unfamiliar to the *S*s.) Hence, the present findings indicate that the effects occur even under conditions that attempt to avoid any suggestion of inconsistency.

The *S*s' level of awareness of inconsistency was not systematically varied within the present experiment, but it is interesting to speculate on the results of such a manipulation. Increasing the person's awareness of his inconsistency by, say, explicitly pointing it out to him might produce greater changes towards consistency. On the other hand, such a procedure could well have the opposite effect by provoking the person's hostility or defensiveness, and so result, not in a greater change towards consistency, but in some other reaction such as the rejection of consistency, with Carlyle, as the hob-goblin of little minds; or in the repression of some of the conflicting issues.

SUMMARY

Two sets of hypotheses were derived from the postulate that people tend to maintain consistency among their opinions on logically related issues. First, it was hypothesized that opinion change would be effected by the Socratic method of simply asking the person to state contiguously his opinions on logically related issues, thus bringing the tendency towards consistency to bear on any logical discrepancies among these opinions. Secondly, it was hypothesized that a persuasive communication directed at a person's opinion on some explicit issue tends to change also his opinions on logically related (even if unmentioned) derivative issues in a consistent direction. A procedure for quantitatively measuring the degree of logical consistency among opinions on related issues was described.

The *S*s were 120 college students participating in three experimental sessions. In the first, before-communication opinions regarding the subjective probabilities and desirabilities of 16 sets of syllogistically related propositions were obtained. In the second session, 16 messages, each arguing for the truth of one premise of each syllogism, were communicated, and the immediate postcommunication opinion of the truth of each of the 48 propositions was obtained. Each *S* served in communication conditions with respect to some issues, and control, no-communication conditions with respect to others. In the third session, one week later, the delayed postcommunication opinions on the truth of the 48 propositions were obtained.

With respect to the Socratic effect hypothesis a significant amount of inconsistency attributable to wishful thinking was found, as postulated on the first elicitation of the opinions ($p < .01$). One week later, on a second elicitation of the opinions, there was a significant ($p = .02$) decline in the amount of this wishful thinking, indicating that the Socratic method did indeed produce the predicted change towards consistency. Several propositional variables were examined for possible relationship to the amount of change towards consistency. It was found that propositions with the predicate term (namely, major premises and conclusions) contributed significantly more ($p = .03$) of the change towards consistency than those without the predicate term.

With respect to the hypothesis concerning logical repercussions, it was found that immediately after the communication there was an induced opinion gain not only on the explicit target issues ($p < .001$) but also the derived issues ($p = .02$), even though neither the derived issues themselves nor even the auxilliary premises needed for their derivation were mentioned in the communication. This

indirectly induced gain on the derived issues fell short, however, of the full amount needed for complete logical consistency ($p < .05$). One week later, there was some suggestion that the impact of the persuasive message had continued to seep down to the derived issues: immediately after the messages the opinions on the derived issues had shown only 52% of the logically required amount of change, while by one week after, this impact had risen to 91% of the required amount ($p = .10$).

REFERENCES

ASCH, S. E. *Social psychology*. New York: Prentice-Hall, 1952.

BAGLEY, W. C. *The educative process*. N. Y.: Macmillan, 1906.

COOPER, EUNICE, & DINERMAN, HELEN. Analysis of the film "Don't be a Sucker": A study in communication. *Publ. opin. Quart.*, 1951, **15**, 243–264.

FESTINGER, L. *A theory of cognitive dissonance*. Evanston: Roe, Peterson, 1957.

HEIDER, F. Attitudes and cognitive organization. *J. Psychol.*, 1946, **21**, 107–112.

HEIDER, F. *The psychology of interpersonal relations*. New York: Wiley, 1958.

HOVLAND, C. I., LUMSDAINE, A. A., & SHEFFIELD, F. D. *Experiments on mass communication*. Princeton: Princeton Univer. Press, 1949.

HOVLAND, C. I., & MANDELL, W. An experimental comparison of conclusion—drawing by the communicator and by the audience. *J. abnorm. soc. Psychol.*, 1952, **47**, 581–588.

JOWETT, B. (Ed.) *The dialogues of Plato*. New York: Random House, 1937.

KATZ, D., STOTLAND, E., PATCHEN, M., & JOCHEN J. Research in methods for changing attitudes. I. Situational and personality determinants of the reduction of prejudice through self-insight. Univer. of Michigan, 1957. (Mimeo.)

KELMAN, H. C. Three processes of acceptance of social influence: Compliance, identification and internalization. *Amer. Psychol.*, 1956, **11**, 261. (Abstract)

OSGOOD, C., & TANNENBAUM, P. The prinicple of congruity and the production of attitude change. *Psychol. Rev.*, 1955, **62**, 42–55.

TANNENBAUM, P. Initial attitude towards source and concept as factors in attitude change through communication. *Publ. opin. Quart.*, 1956, **20**, 413–425.

ZAJONC, R. B. Cognitive structure and cognitive tuning. Unpublished doctoral dissertation, Univer. Michigan, 1954.

(Received March 12, 1959)

Journal of Abnormal and Social Psychology
Vol. 60, No. 3, 1960

DIRECT AND INDIRECT PERSUASIVE EFFECTS OF DISSONANCE-PRODUCING MESSAGES

WILLIAM J. McGUIRE

University of Illinois

THE present study was designed to test some applications of Festinger's dissonance theory (1957) to the area of persuasion. Festinger has predicted (1957, Ch. 9) that a person tends to resist a persuasive message to the extent that it tends to increase dissonance. The present study tests a somewhat broader hypothesis. When a person is presented with a message that argues for a position on an explicit issue which tends to bring his opinion on that issue into conflict with other, unmentioned but logically related issues, he can avoid increased dissonance in two ways: (*a*) by resisting change on the explicit issue and (*b*) by changing also on the related (unmentioned) issues. Hence, it is predicted that a dissonance-increasing message, as compared with a dissonance-reducing one, produces (*a*) less change on the explicit issue, and (*b*) more change on unmentioned but logically related ones.

Furthermore, on the assumption that changes induced in a consistency-increasing direction are stabilized by the bolstering effect of the consonant cognitions, it is hypothesized that dissonance-reducing changes persist longer after the presentation of the communication. Hence, it is predicted that both of the consonant types of change—namely, the changes on the explicit issues produced by the consistency-increasing messages and those on the related, unmentioned issues produced by the consistency-decreasing messages— show slower rates of temporal decay than do the corresponding dissonant changes. Inconsistencies were identified on the basis of an assumed tendency toward wishful thinking.

METHOD

Procedure. There were three experimental sessions held at one week intervals during regularly scheduled classes of a freshman English course. Until the post-experimental "catharsis," the study was represented as an investigation of the relation between reading comprehension ability and the readers' beliefs regarding the material read. At each session a questionnaire was administered to measure the *Ss*' opinions on the truth of 24 critical propositions. At the first session the use of the opinion scales for the truth judgments and also for measuring the desirabilities of the propositions was explained.

At the beginning of the second session the persuasive messages were communicated and memory questions were asked about the content (to support the fiction that we were studying reading comprehension), and then the opinion questionnaire was administered to determine the immediate effects of the communications. In the third session, the questionnaire was again administered (to measure the persistence of the induced opinion changes) and the true nature of the experiment was then explained to the *Ss*.

Material. The opinion questionnaire consisted of a list of 28 propositions. Of these, 24 constituted 8 syllogisms[1] of the "Barbara" type (the other four were repetitions of propositions to provide a reliability check). Beneath each proposition was a scale, described elsewhere (McGuire, 1960), on which the *S* was instructed to indicate his subjective probability regarding the statement's truth. Opinions regarding the desirability of each proposition were obtained on a five-point desirabiltiy scale described by McGuire (1960).

Each persuasive message contained a reasoned appeal, about 250 words long, attributed to a (fictitious) person with a prestige affiliation, arguing in favor of the truth of the premise of one of the eight syllogisms.

Subjects and design. Each of the 35 college freshmen participating in this study received persuasive messages on four of the issues and served as a no-communication control on the other four issues.

RESULTS

Identification of the initial inconsistencies

In order to test the present hypotheses, it was necessary to determine for each message whether it was arguing in direction that increased inconsistency or decreased it. Since each message was aimed at raising the rated truth of a premise, this determination required knowing whether the initial truth rating of the

[1] For a detailed account of the syllogisms and the persuasive messages used in this study, see the ADI document No. 6235 referred to in McGuire (1960) Fn. 1. An example of a syllogism whose premises are rated less desirable than the conclusions (Syllogism 8 in Table 1) is the following: Even should there be a war with Russia, any American city hard to reach by air from the European continent will escape nuclear bomb attack. Our city is hard to reach by air from the European continent. Our city will escape nuclear bomb attack even should there be a war with Russia.

premise had been inconsistently high or inconsistently low on the first administration. To this end, we postulated a tendency toward wishful thinking, amply demonstrated in previous studies by Cantril (1938) and McGregor (1938), such that the rated truth values of propositions are distorted in the direction of their rated desirabilities. The eight syllogisms were therefore divided into four, the premises of which were rated highly desirable relative to the conclusions and four with premises rated relatively low in desirability. In the case of the first four, the persuasive messages were interpreted as arguing in an inconsistency-increasing direction, and for the second four, in a relatively inconsistency-reducing direction.

The data from the first session tend to confirm this auxiliary hypothesis concerning wishful thinking. If the person were being perfectly consistent, his subjective probability of the conclusion should have been directly proportional to the product of the subjective probabilities of the premises (assuming that internal consistency among subjective probabilities requires their following the axioms of probability theory, and that the proportional contribution of the premises to the total probability of the conclusions remained constant from syllogism to syllogism). To the extent that, on the contrary, his beliefs were being distorted by wishful thinking, the excess probability of the conclusions, over the product of the probabilities of the premises, should have been greater in Syllogisms 5 to 8, than in Syllogisms 1 to 4. The data (see Table 1) give evidence of a considerable amount of such distortion. For Syllogisms 1 to 4, the mean excess probability of the conclusions was only 8.99, while for Syllogisms 4 to 8 it was 18.90 (throughout this report, the subjective probability scores cited are the obtained probability ratings multiplied by 100, in order to avoid confusion with the statistical significance levels). The obtained difference of 9.91 is significant at the .05 level. (All significance levels in this report are based on two-tailed analyses of variance tests.)

Wishful and logical thinking

While it was not a primary purpose of the present experiment to compare these two tendencies, the data make possible an estimate of the extent to which wishful thinking and logical thinking each affects opinions regarding the truth of the conclusions. If a person's opinions are to be logically related, then his rating of the probability of the conclusion ($p(c)$) should equal the joint rated probability of the premises ($p(a)p(b)$) plus the rated probability of the conclusion on the basis of all circumstances other than those specified in the premises. Therefore, if we assume that these other circumstances contribute about equal proportions of the total probability of the conclusion from syllogism to syllogism, then consistency requires that the probability of a conclusion be proportional to the joint probability of the premises from which it follows. The group data indicate that, by this criterion there is a high degree of consistency: the correlation between the initial $p(c)$ and $p(a) \cdot p(b)$ for the eight syllogisms is .74.

Wishful thinking accounts for a large part

TABLE 1

TRUTH RATINGS ON THE PREMISES AND THE CONCLUSIONS BEFORE AND AFTER MESSAGES WHICH ARGUED FOR THE INCREASED TRUTH OF THE PREMISES AND LEFT THE CONCLUSIONS UNMENTIONED

Syllogism with Relatively Desirable Premises						Syllogism with Relatively Undesirable Premises					
Syllogism	Excess prob. concl. before	Change immediately after		Further change one week after		Syllogism	Excess prob. concl. before	Change immediately after		Further change one week after	
		Premise	Concl.	Premise	Concl.			Premise	Concl.	Premise	Concl.
1	14.99	18.56	14.63	−12.97	−23.38	5	10.07	18.56	9.64	−13.86	1.81
2	16.14	11.07	4.29	−4.29	−4.64	6	14.99	27.13	12.14	−9.58	−11.66
3	11.00	7.14	12.14	−3.15	2.79	7	20.85	26.00	0.35	−14.99	−5.35
4	−6.28	20.71	13.56	−10.00	−1.78	8	29.49	22.85	−15.35	−18.92	−1.07
Mean	8.99	14.37	11.15	−7.60	−6.75	Mean	18.90	23.65	1.70	−14.34	−4.07

Note.—Scores are given in terms of subjective probability of truth multiplied by 100.

of the remaining variance among the conclusions. When the effects of the desirability ratings of the premises and conclusions are removed by partial correlation, the correlation between $p(c)$ and $p(a)p(b)$ becomes .96, which is significant at the .01 level despite the small number of degrees of freedom.

Since these and the previously reported correlation coefficients are computed from group means it is not legitimate to generalize to the individual case. However, even on the basis of group data it is interesting that 90% of the variance among the mean opinions on the conclusions can be accounted for with just two variables.

Effects Produced by the Persuasive Messages

Opinion changes on the explicit issues. The first experimental hypothesis stated that the persuasive messages would produce opinion change on the explicit issues to the extent they argued in a direction that would increase consistency. Since the auxiliary hypothesis concerning wishful thinking was confirmed, it would be predicted on the basis of the first hypothesis that a greater increase in the $p(a)$ opinions should be produced by the messages directed at the relatively undesirable premises, than by those directed at the relatively desirable.

This hypothesis is supported by the finding that the mean increase in probability on the desirable target premises (of Syllogisms 1 to 4 in Table 1) was 14.37, while the mean for the undesirable premises (5 to 8 in Table 1) was 23.65, a difference significant at the .01 level. That this greater impact on the undesirable premises is not a regression artifact can be demonstrated by further analyses. In the first place, there is not a great deal of difference between the initial probability levels of the two sets of premises. The mean probability rating of the relatively undesirable premises is 50, and of the desirable, 55, which is too small an initial difference to account for the 10-point differential in effect by statistical regression. A still more convincing argument against alternative interpretations is provided by considering the changes from a baseline provided by the control data, that is, changes from before to after the communication for those subjects not receiving a communication on the given issue. From this baseline, the differential is even greater: the mean increase induced by the message on the relatively undesirable premises was 30 and on the desirable, 11. This difference again is significant beyond the .01 level.

Effects of the persuasive messages on the related, unmentioned issues. While some studies report finding that persuasive communications produce little if any effect on related but unmentioned issues (Cooper & Dinerman, 1951; Hovland & Mandell, 1952), McGuire (1960) has demonstrated that remote repercussions in the person's cognitive arena do occur, though usually to a lesser extent than would be required for complete consistency with the induced change on the explicit issue. The present hypothesis states that such remote impact on unmentioned issues is greater for messages arguing in an inconsistency increasing direction, since otherwise the already dissonant opinions on these derived issues would get still further out of line with opinions on the explicit issues. Since all the messages in this study were designed to increase the rated truth of a premise, the prediction follows that the increase of the truth ratings of the conclusion produced by the messages aimed at the desirable premises (Syllogisms 1 to 4) should be greater than those aimed at the undesirable premises (Syllogisms 5 to 8). The prediction is supported by the data. The four conclusions of the desirable premises show a mean increase of 11.15 points after the messages, while the conclusions of the undesirable premises increase by a mean of only 1.70.

This finding is especially striking since the changes induced on the desirable premises themselves were, as has just been described, less than the changes on the undesirable premises. Assuming the subjective probability ratings behave according to the axioms of probability theory, we can compute by substituting the induced change on the premises in the following equation:

$$\Delta P(c) = P(a)\cdot\Delta P(b) + P(b)\cdot\Delta P(a) + \Delta P(a)\cdot\Delta P(b)$$

where $P(a)$ and $P(b)$ are the initial opinions on the premises. Substituting the obtained changes on the premises in this equation, it can be computed that, to be consistent, there

should have been a mean change of 8.30 for the conclusions of the desirable premises and 17.48 for the conclusions of the undesirable. Hence, the obtained changes on the conclusions of the desirable premises exceeded the required amount by 2.85 whereas the obtained change on the conclusions of the undesirable premises fell short of the required amount by 15.78 points, a difference significant at the .01 level. The differential is reduced when, to control for possible regression effects, the changes in the communication conditions are computed from a baseline provided by the control (no communication) conditions, but the reduced difference is still significant at the .05 level.

The relative persistence of consistent and inconsistent opinion changes. On the assumption that the bolstering effect of consonant cognitions would stabilize changes in a direction reducing inconsistency, it was hypothesized that such changes would be more persistent than the dissonance-increasing changes. Hence, we predicted that as regards the changes on the premises themselves, increases on the undesirable premises would be more lasting than those on the desirable; regarding the derived changes on the conclusions, we predicted, conversely, that increases on the conclusions of the desirable premises would be more lasting than the increases on the conclusions of the undesirable premises.

Neither of these predictions is supported by the data. In regard to the persistence of the induced increase on the explicit issues, 7.60 points of the originally induced 14.37 on the desirable premises dissipated during the week; while 14.34 points of the originally induced 23.65 on the undesirable premises dissipated (see Table 1). Hence, in terms of the arithmetic loss during the week, the dissonance-reducing changes show 6.74 points more decay than those which increased dissonance, reversing the prediction that the dissonance-increasing changes would show more decay. In terms of percentage loss of originally induced increase, the two conditions show approximately equal losses at the end of one week. The prediction regarding greater stability of dissonance reducing changes on the remote issues likewise is not confirmed by the data. The relatively desirable conclusions show a 4.07 point decline during the postcommunication week, and the undesirable conclusions, a decline of 6.75 points. Hence, the dissonance-increasing changes are slightly more persistent than the dissonance reducing ones when measured in terms of amount of decay; and the dissonance-reducing changes slightly more persistent when measured in terms of rate of decay.

DISCUSSION

The confirmation of the hypothesis of greater immediate impact of the dissonance-reducing communications is particularly striking because, in the case of both the explicit and the related issues, the tendency toward wishful thinking works against it. The implication of our finding is that a person can more easily be argued into an increased acceptance of the truth of an undesirable state of affairs than a desirable one. The underlying reason for this, suggested by dissonance theory, is that before the persuasive communication, he has already overrated the truth of the desirable outcome and because of his need for consistency, therefore resists further uprating that would make his expectation become even more dissonant with his cognitions of the available evidence.

The confirmation of both derivations from dissonance theory regarding the immediate impact of the persuasive messages but the complete nonsupport for the derivations regarding the persistence of these immediate effects suggests that predictive power of the theory could be improved if a postulate concerning "salience" were added to it. Such a postulate would recognize that all the person's cognitions relevant to a given opinion might not be equally brought to bear on the opinion in a given situation. Rather, one or another subarea within this total cognitive arena might be made more salient by experimental manipulation. It might often happen that the person's cognition on a given issue could be made to appear dissonant in either one direction or the opposite, depending on which of these subsets of relevant cognitions is made salient.

Applying this formulation to the present study which investigates the person's expectations regarding desirable and undesirable states of affair, we can detect two cognitive forces: wishful thinking and logical thinking. The person presumably had initially arrived

at a steady state of expectation on the issues which represented a compromise between the two needs—wish fulfillment and reality orientation—a compromise that was fully consonant with neither. In this study the experimental manipulation involved presenting messages that contained rational arguments regarding the probability of the events in question. We can reasonably interpret this procedure as producing increased saliency of the cognitive realm involving logical thinking, reality orientation and objective evidence. Hence, any underlying dissonance between expectations and evidence would be activated, with a resulting readjustment of the cognitions towards greater consonance with the evidence and less with desires.

After the person leaves the experimental situation, however, the increased saliency of the evidential subarea of his cognitive arena should tend to subside to its pre-experimental level. Consequently, the need for consonance between expectations and desires once again becomes relatively more salient, and the dissonance between these beliefs and hopes is now felt behaviorally to a greater extent. By adding such a postulate of salience, dissonance theory could better fit the present results, both as regards initial impact and persistence of effects.

SUMMARY

Several hypotheses regarding the effectiveness of persuasive messages were derived from Festinger's "Theory of Cognitive Dissonance." It was predicted that messages would be effective in changing opinions on the explicit issues to the extent that they argued in dissonance-reducing directions; and on related, unmentioned issues, to the extent that they argued in a dissonance-increasing direction. It was further predicted that the dissonance-reducing changes would show a slower rate of temporal decay after the communication.

These hypotheses were tested by asking 35 college students their opinions regarding the truth of several syllogistic sets of premises and their valid conclusions. The *S*s were then presented with messages which argued for the increased truth of the premises (leaving the conclusions unmentioned). Immediately after the persuasive messages, and again one week later, the opinions on the truths of both premises and conclusions were again measured.

On the basis of desirability ratings of the propositions and of an assumption concerning wishful thinking, the syllogisms were partitioned into two subsets: those whose premises received inconsistently high truth ratings initially and those receiving relatively low initial truth ratings.

Both predictions regarding the immediate impact of the messages were confirmed. On the explicit issues (the premises), more opinion change was produced by the inconsistency-reducing messages ($p < .01$); on the related, unmentioned issues more change was produced by the inconsistency-increasing messages ($p < .05$). The prediction of greater permanence for the dissonance-reducing changes was not confirmed: the dissonance-reducing and dissonance-increasing changes both showed about the same rate of temporal decay. On the basis of these findings, it was suggested that a postulate concerning "salience" be added to dissonance theory.

REFERENCES

CANTRIL, H. The prediction of social events. *J. abnorm. soc. Psychol.* 1938, **33**, 364–389.

COOPER, EUNICE, & DINERMAN, HELEN. Examination of the film, "Don't be a Sucker": A study in communication. *Publ. opin. Quart.*, 1951, **15**, 243–264.

FESTINGER, L. *A theory of cognitive dissonance.* Evanston: Roe, Peterson, 1957.

HOVLAND, C. I., & MANDELL, W. An experimental comparison of conclusion drawing by the communicator and by the audience. *J. abnorm. soc. Psychol.* 1952, **47**, 581–588.

McGREGOR, D. The major determinants of the prediction of social events. *J. abnorm. soc. Psychol.*, 1938, **33**, 179–204.

McGUIRE, W. J. Cognitive consistency and attitude change *J. abnorm. soc. Psychol.* 1960, **60**, 345–353.

(Received March 9, 1959)

Journal of Abnormal and Social Psychology
Vol. 60, No. 3, 1960

SOME CONSEQUENCES OF EXPERIMENTAL REJECTION

ALBERT PEPITONE AND CHESTER WILPIZESKI[1]

University of Pennsylvania

REJECTION by others has been cited as the determinant of a large number of different responses, pathological and normal. It is widely believed, for example, that social hostility as reflected in delinquent behavior stems from a parental environment in which the child is somehow isolated and unloved (cf. Bandura & Walters, 1958 for a recent variant of this view). Rejection has also been considered an important basis of suicide and other self-punitive responses, as well as of social withdrawal and apathy (cf. Wolberg, 1944). Finally, a laboratory study by Dittes and Kelley (1956) showed that subjects (*Ss*) who were given to believe they had low sociometric status in the group conformed more to the opinions of that group than those who thought they were highly accepted.

It is not clear from these studies whether such diverse responses occur under exactly the same conditions of rejection, and if so, whether they are displayed by the same person. The unanswered question is: what responses can be expected from a given kind and degree of rejection, and theoretically, why?

A number of experiments (Pepitone, 1958) have shown that the magnitude and the negative effects of various "ego-threats" can be reduced when the person perceives certain causal properties relating to responsibility, intentionality, etc. in the threatening situation. On the basis of these results, a theoretical process may be proposed to account for some of the responses reported to be associated with rejection. The assumption is made that rejection from an important group generates more or less severe personal "devaluation anxiety" in most individuals, and that more or less strenuous attempts are made to reduce such anxiety. Anxiety-motivated instrumental behavior may be primarily directed toward changing the social environment. For example, rejected persons may comply or identify with the rejector's opinions, become compulsively friendly, withdraw from the group, or display various other *socially adaptive* responses. However, when socially adaptive responses are unsuccessful, impossible, or when for any reason, the anxiety is irreversible, the person then attempts to make the anxiety more tolerable to himself. In other words, when the rejection or other basis of devaluation anxiety is a *fait accompli*, persons tend to respond in a way that is primarily *self-adaptive* in function.

The basic assumption of the present experiment, suggested by the studies referred to above, is that the person can make irreversible devaluation anxiety more tolerable by explaining or fixing responsibility for it. Specifically, it is hypothesized that "causal attibution" to the two general sources of responsibility is reflected empirically in two kinds of responses:

1. *Self-depreciation.* An "internal" attribution of responsibility for the rejection is shown by some form of self-depreciation on the part of the rejectee. Such a response in effect acknowledges or "accommodates" the personal devaluation implicit or explicit in the rejection and thereby makes it more bearable.

2. *Hostility.* Various forms of aversion toward the social environment may be regarded conceptually as an external attribution of responsibility for the rejection. Thus, by disliking the rejectors, by finding fault with them and the social situation generally, the person "accounts" for the rejection, and palliates the devaluation anxiety generated by it.

An important question is whether the two hypothesized responses to rejection tend to occur in the same person. From one theoretical standpoint, it may be supposed that if in the

[1] This experiment is one of a series supported by the Office of Naval Research through a contract with the senior author N-ONR 551-(27). We gratefully acknowledge this support, although we must take full responsibility for the views expressed in this paper. We have also benefited from a critical reading of the manuscript by J. Aronfreed and S. Feshbach. We wish to acknowledge the help of R. C. Mackowiak in parts of the statistical analysis.

past the person has succeeded in lowering rejection anxiety by internal attribution (self-depreciation), he should be less likely to produce an external attribution (hostility). That is, the two responses should correlate negatively. On the other hand, causal attribution in general may have been "reinforcing," in which case all responses in that class would tend to be positively correlated in individuals. From a different theoretical slant, a positive correlation would result if the relative contributions to the rejection are uncertain. Where the person is not sure of the specific basis for being rejected, he might be expected to implicate both himself and the rejectors. Thus, an unequivocal prediction regarding the relationship among responses to rejection is not possible.

There are of course, all degrees of rejection. On the assumption that stronger rejection would tend to produce more vigorous self-adaptive responses, an attempt was made operationally to establish Implicit and Explicit levels of rejection. As it turned out empirically, the differences in effects between the two conditions were negligible. *S*s under the two conditions were, therefore, combined for purposes of analysis.

METHOD

Ninety-five college undergraduate male volunteers served as *S*s. Each three-person group consisted of one *S* and two confederates.[2] The behavior of the confederates was completely prescribed in advance by a well rehearsed script.

After assembling the group in the laboratory, *E* informed the *S*s that they were going to take part in an important survey concerned with the breakdown of family structure. The purpose of the study was to investigate the attitudes of college undergraduates toward two issues relevant to family disorganization: (*a*) premarital sex relations, and (*b*) drinking among minors. Specifically, *S*s were asked to reflect upon the question: "Do you, in general, encourage or discourage premarital sex relations?" and to jot down the three strongest reasons for their opinion. Lots were then ostensibly drawn to determine who would discuss his views first. However, the drawing was so contrived that *S* always led off. He declared himself either pro or con on the question and then elaborated his views by citing the reasons he held to be decisive. Finally he was requested to raise up in full view one of a series of cards numbered 1 to 6, representing the degree to which he personally either encouraged or discouraged premarital sex. The identical procedure was followed by the two confederates. In the *No Rejection* condition, the confederates held the same opinions as *S*, and assigned approximately the same numerical weight to them as did *S*. In both the *Implicit* and *Explicit* conditions, the confederates differed from *S* by adopting the diametrically opposed opinion. The reasons given by the collaborators and the numerical strength of their opinions were such as to maximize the isolation of *S* while maintaining the naturalness of the situation.

Following the 15-min. discussion, the groups under the rejection conditions were given a brief recess before taking up the second topic. The *Explicit Rejection* procedure took place during this recess. *E* left the room mentioning that he had forgotten some forms. The two confederates then became engaged in friendly conversation. They were instructed to ignore or to respond curtly to any of *S*'s questions or remarks, and to present an unfriendly demeanor when glancing at *S*. A number of specific actions reinforced the personal rejection. For example, one confederate offered a cigarette (if *S* was a smoker) or chewing gum to the other, but pointedly neglected *S*. In the *Implicit Rejection*, there was silence during this period, unless *S* initiated conversation, in which case the responses of the confederates were minimal and neutral.

When *E* returned, he reluctantly announced that time prevented any further discussion of the topic; that it was a closed issue. He added that he would like to have them make a few ratings which would be helpful in interpreting the discussion. Following this rating procedure, *S*s were asked whether they would like to return for a discussion of the second topic with the same group. Assurance was given to each *S* throughout the postdiscussion procedure that his answers would be treated anonymously and confidentially. No *S* would know how the other responded. After collection of these data, *E* informed the *S* of the nature of the experiment, exposed the dissimulation, and requested that he refrain from discussing it with others.

RESULTS

Adequacy of Experimental Operations

Two questionnaire items may be examined for evidence that the manipulations had the desired effects. In one case, *S* is asked to make a quantitative rating of the degree he believes he is liked by the other *S*s. Responses were recorded on a 20-unit scale ranging from "dislike me very much" on the low end to "likes me very much" on the high. The second item requests a rating of how valid he thinks the other *S*s regard his opinions as expressed in the discussion. Responses on this 20-unit scale

[2] There were actually four individuals who played the two confederate roles approximately the same number of times in each experimental condition.

could range from "highly invalid" on the low end to "highly valid" on the high.

Table 1 contains the mean ratings of the two items under the experimental conditions. It is clear that the experimental treatments had a significant effect on both measures. Rejected *S*s felt less well-liked personally and felt that their opinions were thought to be less valid than nonrejected *S*s. It can also be seen that while the means drop regularly across conditions, consistent with the assumption of a quantitative continuum of rejection, the differences between the Implicit and Explicit Rejection conditions are not statistically significant. Subsequent analysis will deal, consequently, with the Implicit and Explicit *S*s combined into a single Rejection condition.

TABLE 1

MEAN RATINGS ON MEASURES OF REJECTION OPERATIONS

	No Rejection ($N = 32$)		Implicit ($N = 32$)		Explicit ($N = 31$)	F	p (one-tail)
S's perception of how others[a] evaluate his opinions	20.4	$t = 2.68$, $p < .01$	16.6	*ns*	14.6	8.45	<.001
S's perception of how others[b] like him personally	14.7	$t = 7.84$, $p < .01$	7.1	*ns*	6.8	41.9	<.001

[a] Ratings of each confederate combined.
[b] Ratings of each confederate combined.

TABLE 2

MEAN RATINGS ON MEASURES OF SELF-DEPRECIATION

	No Rejection ($N = 21$)	Rejection ($N = 41$)	t	p (one-tail)
Positive minus negative personal self-evaluations	+3.1	+1.7	1.97	<.03
S's evaluation of his own opinions	($N = 32$) 16.8	($N = 63$) 14.9	3.13	<.005

Self-Depreciation

As hypothesized above, an irreversible rejection can be made more acceptable by taking the blame for it,' that is, by lowering the self-evaluation in some degree in some circumscribed area. Two sources of data were assumed to be relevant to the hypothesis. One is a listing of positive and negative personality characteristics that all but a few *S*s provided following the questionnaire. The specific measure of this "personal" self-evaluation consists of the number of positive attributes minus the number of negative ones. The second measure of Self-depreciation is based on a quantitative rating by *S* of the validity of his opinions expressed in the discussion. Responses were recorded on a 20-unit scale.

Table 2 contains the average discrepancy between *S*'s positive and negative self-evaluations, and the mean ratings of the validity of his opinions under the two experimental conditions. It is evident that the average number of *S*'s positive self-evaluations as well as the estimate of the validity of his opinions were lower in the Rejection than in the No Rejection condition. Thus, whether the evaluation of himself is in personal terms or in terms of the worth of the opinions he holds, rejection definitely produces Self-depreciation.

Hostility

An external attribution of responsibility hypothesized to follow from rejection can be reflected in many forms of social aversion or hostility. Several measures were developed for different aspects of this general response tendency. To get at personal hostility, *S*s were asked to rate how much they liked or disliked each of the two confederates. A more impersonal measure of hostility called for *S* to rate the validity of each of the confederate's opinions. As an indirect measure of social hostility, *S* rated his amount of interest in the discussion issue. All of these ratings were on 20-unit scales. Finally, the most general measure of hostility was a quasi-behavioral one—a statement of whether or not *S* would commit himself to return for another discussion with the same discussion partners.

Table 3 shows the mean ratings on the three scales and the percentage of *S*s who express a willingness to return, under the two experi-

mental conditions. All four measures assumed to represent hostility were significantly affected by rejection. Not only were the rejectors' opinions held to be less valid, but personally the rejectors were more disliked than the discussion partners were liked by *S*s in the No Rejection condition. Moreover, *S*s who had been rejected found the discussion topic less interesting than the *S*s in the No Rejection condition. Finally, for whatever reason, many more rejected *S*s were averse to returning for another discussion. In all, there is little doubt that the *S*s who were implicitly or explicitly rejected find their social environment relatively unattractive.

We turn now to a consideration of whether the responses produced by the experimental conditions are in any way related to each other within the same individual. Table 4 contains the product moment coefficients of correlation among the responses under each experimental condition. With respect first to the Rejection condition, there are two trends worth noting: (*a*) self-evaluation, in terms of the validity of one's opinions, tends to be positively correlated with each of the three measures of Hostility, and (*b*) liking others is positively correlated with one other measure of Hostility, namely, the evaluation of the opinions others hold.

Of greatest theoretical interest are the significant positive relationships between one of the Self-depreciation measures and two measures of Hostility. To determine whether the experimental rejection had anything to do with this result, the correlations under the No Rejection condition may be examined. It will be noted that the correlations between Validity of Own Opinion and each of the three Hostility responses, like their counterparts under the Rejection condition, are positive (two of the coefficients are at least at the .10 level). It thus seems fairly clear that a positive relationship between Self-depreciation and Hostility does exist for some measures, but is not due to the effects of experimental rejection.

In fact, when the differences between the correlation coefficients are tested for significance (by way of the *z* transformation and the *SE* of the difference between *z* scores), none of them is significant. In other words, no correlation seems to have been affected by the experimental rejection. Thus, the relationship between liking others and the evaluation of their opinions is significantly positive under the No Rejection condition as it was under the Rejection condition. It should also be noted that although the difference between experimental conditions falls short of significance, two of the correlations in which

TABLE 3

MEASURES OF VARIOUS FORMS OF HOSTILITY

	No Rejection (N = 32)	Rejection (N = 63)	t	p (one-tail)
Personal liking of others'[a]	21.3	18.3	2.21	<.03
S's evaluation of others' opinions[b]	28.4	17.6	4.76	<.001
S's interest is discussion	15.4	13.8	1.96	<.03
S's commitment to return	69%	50%	4.11[c]	<.03

[a] Ratings of the two confederates combined.
[b] Ratings of the two confederates combined.
[c] Chi square with 1 *df*.

TABLE 4

CORRELATIONS AMONG RESPONSES PRODUCED UNDER REJECTION AND NO REJECTION CONDITIONS

	Valid own Opinion		Like Others		Valid Others		Interest	
	No Rej.	Rej.	No Rej.	Rej.	No Rej.	Rej.	No Rej.	Rej.
Personal Self-Eval.[a]	+.39	−.09	−.02	−.05	−.03	−.15	+.39	−.05
Valid own opinion			+.36	+.46*	+.26	+.18	+.43*	+.28*
Like others					+.43*	+.37*	−.12	+.21
Valid other opinion							+.12	+.03

* These coefficients are at least of the size required for the .05 level of significance based on both tails of the *t* distribution.
[a] Note that correlations involving this measure are based on *N* of 21 and 41 in the No Rejection and Rejection conditions respectively. All other measures are based on *N*s of 32 and 63.

Personal Self-evaluation is involved tend to be strongly positive ($p < .10$) under the No Rejection condition. There is no ready explanation of this, assuming that it is not merely a chance pattern.

DISCUSSION

There seems little doubt that rejection produces an increase in the strength of both self-depreciation and hostility in the sample of *S*s employed. Such findings are consistent with the hypothesis that an unchangeable rejection tends to bring about self-adaptive attributions of causality. Since, however, the experiment cannot prove the hypothesis unequivocally, it becomes necessary to summarize the arguments that support the hypothesis and oppose alternative interpretations.

Considering the finality of the rejection, the clear impossibility of communicating with the rejectors, and the anonymity of the responses it may be assumed that the behaviors produced by rejection did not primarily have a socially adaptive function. Thus, concretely, it appears to us to be inadequate to interpret the hostility merely as inhibited aggression toward the rejectors. Similarly, under the circumstances, self-depreciation could hardly have been produced in order to gain the rejectors' sympathy and thereby to rectify the damage to the self. On the other hand, it seems more plausible to suppose that the manifestly irreversible rejection strongly inclined the *S* to adjust himself to the negative situation rather than to try to change the situation. Although presumably there are other kinds of self-adaptive responses, hostility and self-depreciation appear to represent attributions of responsibility most closely. Clearly more definitive criteria of self-adaptive causal attributions are needed for further confirmation and extension of the theory.

It is important to point out that there were sources of devaluation anxiety in the experiment in addition to the rejection itself. For example, the discussion topic of premarital sex relations undoubtedly was a source of strong anxiety in many *S*s under both experimental conditions.[3] It is quite probable that the shame engendered in this way produced self-depreciation and even some hostility. In addition, there is the possibility that the very articulateness and urbanity of the confederates in agreeing and disagreeing with the *S*s made some of the latter feel extremely inadequate. This state of affairs might very well have created anger toward the confederates and attendant guilt, which in turn could have produced some self depreciation. Here again, however, granting our "rationalistic" classification of responses into socially and self-adaptive categories, there would not seem to be much of a basis for such "retaliatory" anger.

Given the motivation to make the self-adaptive responses of hostility and self-depreciation in both experimental conditions, what accounts for the fact that seemingly alternative self-adaptive responses tend in some cases to be positively correlated? One compelling possibility may be mentioned. In the concrete experimental situation of the *S*, it was not at all clear who was responsible for the rejection. It may be supposed that when the causal situation is basically uncertain, persons who are motivated to make self-adaptive responses tend to "spread" the responsibility over the major possible causes. Such a tendency would make for positive correlations among the responses.

[3] Concerning this topic, the breakdown "For" and "Against" was approximately the same in all experimental conditions.

SUMMARY

The experiment was designed to test the hypothesis that irreversible rejection gives rise to attributions of responsibility for that rejection, toward the self (self-depreciation) and toward the rejectors (hostility). The theoretical basis for the hypothesis was that personal devaluation anxiety engendered by the rejection can be made less unacceptable to the self through a causal analysis of it. The data were consistent with the hypothesis. A correlational analysis disclosed that a positive correlation existed between one of the self-depreciation responses and two hostility responses under both conditions of Rejection and No Rejection. The discussion pointed to sources of devaluation anxiety other than rejection, and to the supposed tendency for the person to spread responsibility in a situation

where the cause of the anxiety is relatively unclear.

REFERENCES

BANDURA, A., & WALTERS, R. Dependency conflicts in aggressive delinquents, *J. soc. Issues.* 1958, **3**, 52–65.

DITTES, J., & KELLEY, H. Effects of different conditions of acceptance upon conformity to group norms, *J. abnorm. soc. Psychol.*, 1956, **53**, 100–107.

PEPITONE, A. Attributions of causality, social attitudes and cognitive matching processes. In R. Taguiri & L. Petrullo (Eds.), *Person perception and interpersonal behavior*, Stanford: Stanford Univer. Press, 1958.

WOLBERG, L. The character structure of the rejected child. *Nerv. Child*, 1944, **3**, 74–88.

(Received April 1, 1959)

Journal of Abnormal and Social Psychology
Vol. 60, No. 3, 1960

STIMULUS POOLING AND SOCIAL PERCEPTION[1]

RICHARD H. WILLIS[2]

University of Wisconsin

THE purpose of this experiment was to study the effect of stimulus pooling on judgments of personal attractiveness, where, in this context, stimulus pooling refers to the assembling of stimulus components into composite stimuli. An understanding of the relations between judgments made to composite stimuli and judgments made to components presented alone is of fundamental importance for the psychology of judgment. It has, in addition, important implications for social perception, e.g., for the development of social stereotypes and the formation of impressions of others.

Very briefly, the method employed was first to scale identification photographs on personal attractiveness, and then to assemble these photographs into sets of two or three. These sets, or composite stimuli, were then judged for overall attractiveness by additional judges.

One particularly simple kind of relationship that might be expected to hold between measures of judgments made to composite stimuli and those made to the stimulus components presented alone is that the former are equal either to the mean or the sum of the latter, depending upon the kind of task set for judges. Following Gulliksen (1956), this "linearity" or "additivity" of pooling effects may be derived by starting with the assumption that the rate of change of subjective value, v, with respect to an objective stimulus measure, x, is constant. This corresponds to the differential equation

$$dv/dx = k \qquad [1]$$

Upon integration and setting $x = x_0$ when $v = 0$, we obtain

$$v = k(x - x_0) \qquad [2]$$

For stimulus components i and j, and composite stimuli ij,

$$v_i = k(x_i - x_0) \qquad [3]$$

$$v_j = k(x_j - x_0) \qquad [4]$$

$$v_{ij} = k(x_i + x_j - x_0) \qquad [5]$$

By eliminating x_i and x_j between Equations 3–5, we have

$$v_{ij} = v_i + v_j + kx_0 \qquad [6]$$

Since scaling data give, not the subjective values, v, but scale values, s, which differ from the former by a constant, we substitute $s_i - C = v_i$, $s_j - C = v_j$, and $s_{ij} - C = v_{ij}$. The resulting expression is

$$s_{ij} = s_i + s_j + (kx_0 - C) \qquad [7]$$

Because judges in the present experiment were requested to respond to the *mean* of component scale values, rather than to their sum, we rewrite Gulliksen's equation (1956) as follows

$$s_{ij} = \left(\frac{s_i + s_j}{2}\right) + K \qquad [8]$$

This indicates that a plot of the scale values of composite stimuli against mean scale values of stimulus components will be linear with intercept $K = \frac{1}{2}(kx_0 - C)$ and—note—unit slope. When $kx_0 - C = 0$ in Equation 7, or $K = 0$ in Equation 8, we may speak of *additivity*, whereas the term *linearity* may be reserved for the more general relationship.

Gulliksen (1956) working with food preferences and using the method of pair comparisons, found Equation 7 to fit his data as well as three more complex functions, viz. logarith-

[1] This paper is based on a dissertation submitted to the graduate school of the University of Wisconsin in partial fulfillment of the requirements for the PhD degree. The writer wishes to express his sincere appreciation to J. C. Gilchrist, under whose supervision this investigation was conducted, and to David A. Grant and Leonard Berkowitz, who served as members of the author's doctoral research committee.

[2] Now at the Carnegie Institute of Technology.

mic, square root, and negative exponential. Rimoldi (1956) working with "interest 'in knowing' the person or persons" and using three different methods of scaling, found correlations between obtained scale values of composite stimuli and those predicted from an additive relationship (Equation 8 with $K = 0$) to be .93, .97, and .96.

On the other hand, what is known about the nature of social stereotyping suggests that under certain conditions, judgments made about composite stimuli (social groups) may not be so simply related to the mean judgments made about stimulus components (individual members of the groups involved).[3]

Campbell (1956), for example, has derived a principle of "enhancement of contrast" from the concept of stimulus generalization, whereby the prediction is made that the extent of overlap between two groups will be underestimated at intermediate levels of learning, due to a reduction of the variances of judgments within a group. Generalization among members of a group causes those relatively high on the continuum to be underestimated, while those relatively low are overestimated. This prediction was verified experimentally, using nonsense syllables. Clarke and Campbell (1955) found a slight degree of such relative exaggeration of group differences for white students' estimates of the examination scores of their Negro and white classmates.

However, Campbell's derivation applies to judgments of individual group members, not to judgments of the groups as such. A study which demonstrates the operation of enhancement of contrast in judgments of groups is that of Kephart (1954). He found a marked tendency for policemen to exaggerate the Negro-white crime rate differential in their districts. Interestingly, the higher the actual arrest rate in the district, the greater the tendency towards overestimation. This fact by itself does not imply that pooling effects are necessarily nonlinear, but it does imply that Equations 7 and 8 are not sufficiently general. In order to obtain the kind of enhancement of contrast found by Kephart, a slope greater than unity is required.

So far as the author is aware, no prior investigation has made a direct comparison of judgments of composite stimuli and those of stimulus components and at the same time has employed stimulus material somewhat analogous to that upon which judgments of social groups and their members is based. The main question to which the reported research was directed may be formulated as follows:

Are the processes involved in responding to composite stimuli by themselves sufficient to yield the exaggeration of group differences so often found in social sterotypes?

In the absence of any precise knowledge of the weighting laws involved, it seemed desirable to include as independent variables, in addition to differences in means of composite stimuli, such "intragroup" factors as heterogeneity and skewness of component scale values about composite stimulus means. Because interest was centered about pooling effects, it seemed obligatory to include as an independent variable the number of components per composite stimulus. Presumably any effects due to stimulus pooling, such as enhancement of contrast, should be accentuated by increasing the number of components pooled. Certain other variables, such as sex of judges, sex of photographic subjects, and order of presentation, were introduced primarily as experimental controls.

And finally, one additional variable, the number of judges working together, was included for the purpose of testing a specific hypothesis. It was expected that pairs of judges working together, making a single, collective response to each stimulus, would produce distributions of judgments with less dispersion than those produced by individual judges. It was hypothesized that a process of social accomodation would occur, whereby partners would reconcile differences in judg-

[3] The situation is complicated—as, for example, Allport (1954) has pointed out—by the fact that social stereotypes often derive, not only from contact with group members, but also in large part from attitudes assimilated from nonmembers, especially those belonging to the individual's reference groups. Furthermore, systematic distortion of memory traces, so well documented by Bartlett (1932), plays an important role in the development of stereotypes. An analysis of such effects, however, is outside the scope of the present paper.

ments by compromising, thereby avoiding extreme judgments.

METHOD

Subjects. The *S*s were 212 University of Wisconsin undergraduates recruited from elementary psychology classes during the academic year 1956–57. Thirty-two *S*s served as judges during the scaling of the stimulus components, while 180 *S*s participated in the judging of the composite stimuli.

Scaling the stimulus components. Approximately 1,500 photographs, 1¾ in. square, of students entering the University in the fall of 1950, were obtained. From this initial group, two subgroups of 200 each were selected of males and females. Each subgroup included relatively high proportions of extremely attractive and extremely unattractive individuals.

Values on a scale of personal attractiveness were obtained for each of the 400 photographs by means of the method of equal-appearing intervals. The judges, 16 males and 16 females, used a sorting rack containing nine categories labelled A through I. A front cover on the rack prevented judges from observing the number of photographs placed in each category, thus increasing the independence of judgments. No restrictions were placed on the number of photographs to be placed in each category. Instructions (cf. Willis, 1957) were read by the judge, and then *E* explained the main points of the task. One of the two groups of 200 photographs was sorted first, then the other. Order of presentation was balanced both for male and female judges.

Experimental design: Overview. The entire experiment, subsequent to the scaling of components, may be considered and analyzed either as three $2^3 \times 4$ orthogonal factorials,[4] or as two interlocking $2^4 \times 4$ orthogonal factorials. In all, there were eight factors, of which seven had two levels and one (heterogeneity) had four.

Assembling the composite stimuli. Once the scale values of attractiveness were obtained, photographs were assembled into composite stimuli, or sets, and mounted on 3×5 index cards. Each set consisted of either two or three photographs (2-sets and 3-sets, respectively). Sets were systematically varied with respect to the following variables: (*a*) *set size*—the number of photographs per set, (*b*) set *mean scale value* (MSV)—the mean attractiveness scale value of set components, (*c*) *heterogeneity*—mean deviation of scale values about MSV, (*d*) absence or presence of *skewness* of the distribution of scale values about MSV, and (*e*) *sex of photographic subjects*, all photographs in a set being of the same sex.

Assembling the subseries and series. A number of sets were assembled into a *subseries*, a subseries being the sets given to a judge to sort at one time. All sets within a subseries were of the same size and sex. Each judge was required to sort matching male and female subseries, and order of presentation was balanced for each group of judges. There were three pairs of matching subseries, each pair consisting of a male and a female subseries. Each pair of matching subseries formed a *series*. Series I contained 2-sets. Series II contained symmetrical 3-sets, and Series III contained highly skewed 3-sets, where the measure of skewness was twice the intermediate scale value minus the sum of the other two. Within each MSV condition, skewness was always in the same direction; in the high MSV condition, the most extreme component in each set was also the one with the highest scale value, whereas the most extreme component in each low MSV set was the one with the lowest scale value.

Within each series, sixteen *set types* were represented. These set types were formed by the various combinations of male and female photographic subjects, high and low MSV, and four levels of set heterogeneity. Each set type in Series I was represented by six sets. In Series II and III, each set type was represented by four sets, with the exception of two set types which were represented by two sets. The six subseries were all matched, set type by set type, on both MSV and heterogeneity. In general, it was possible to achieve excellent matching.

Obtaining the judgments of composite stimuli. Sixty judges were used with each series, each judge being used with only one series. In half the sortings, a single judge responded to each set, using the same sorting rack used in the scaling. In the other half of the sortings, two judges of the same sex worked together in an informal way, making decisions about the placement of sets into the sorting rack. Judges working alone read a set of instructions which were identical with that used in the scaling, except for the addition of an instruction to take all photographs of each set into account in making judgments. Instructions read by partners were identical to those read by set judges working alone, except for an additional item instructing partners to alternate in the actual handling of sets, while deciding together about the placement of each set. *E* reviewed the important points in the instructions for judges before they began, just as was done during the scaling.

Sixteen scores were obtained from each *J*,[5] one for each set type. The score for a set type was the mean score for all sets of that type, while the score for a set was merely the number of the response category into which it was placed, categories being numbered 0 to 8, from least to most attractive.

[4] There were two error terms, Error Term I (MS, *J*s within conditions), and Error Term II (MS, residual). Error Term I was used to test the significance of differences between independent groups, while Error Term II was used with tests based on possibly correlated data, i.e., tests of sources of variation involving one or more of the factors of heterogeneity, MSV, and sex of photographic subjects—and thus also involving repeated measures. The *df* for Error Term I was equal to 32 times the number of series involved, while the *df* for Error Term II was equal to 480 times the number of series involved.

[5] The word *judge* will hereafter be used to designate the individual, either working alone or with a partner, while the symbol *J* will be reserved to refer to the judging unit, which may consist of either one person or two. A *J* corresponds to a subject for purposes of statistical analysis.

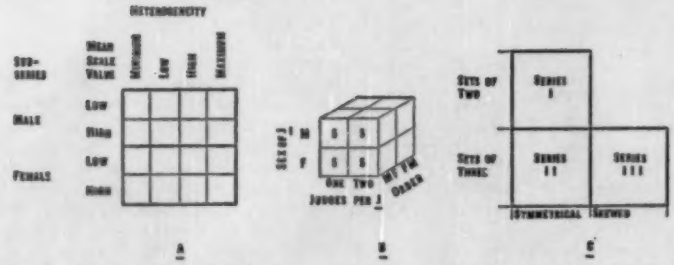

FIG. 1. The design of the experiment: the sixteen set types within a series, the eight subgroups of *J*s within a series, and the relationships between series with respect to set size and presence of skewness within sets.

Experimental design: Conditions. For each series, *J*s were balanced with respect to sex and number of judges per *J*. Order of presentation of subseries was also balanced. The entire design, which is diagrammed in Fig. 1, was replicated five times. In Fig. 1A are shown the sixteen set types that occurred in each series. There were eight experimental conditions for each series, each containing five *J*s, as shown in Fig. 1B, making 120 *J*s in all. Fig. 1C diagrams the interrelations between series. Series I and II are both free from skewness within sets, but differ in set size. Series II and III have set size in common, but the sets in Series III are skewed to the full extent allowed by the specified heterogeneity. Series I and III thus differ both in size of set and presence of skewness. All three series were matched on each of the six factors shown in Figs. 1A and 1B. Each series, taken separately, forms an orthogonal design in these six factors. Series I and II taken together form an orthogonal design in these six factors plus set size, while Series II and III taken together form an orthogonal design in these six factors plus skewness. Heterogeneity, MSV, and sex of photographic subject are within-*J* factors; number of judges per *J*, sex of judges, and order of presentation exist between conditions within series; and set size and skewness are between-series factors.

RESULTS

Overall main effects. In Table 1 are shown the subgroup means obtained when data from all three series[6] were subdivided by levels of each factor in turn. The results of the tests of significance are reported in Table 2 (See Footnote 4). Heterogeneity, set MSV, and sex of photographic subjects, tested against Error Term II, are all highly significant. None of the remaining overall main effects approach significance.

[6] There are two exceptions: the set size and the skewness comparisons were based on data from two series only (Series I and II, and Series II and III, respectively), to avoid confounding the set size and skewness effects within a comparison. See Fig. 1C.

Set heterogeneity in moderate degree showed no effect, but mean set judgments were slightly lowered in the condition of maximum heterogeneity. A Duncan Range Test (Duncan, 1955) indicated that the three highest means were not significantly different from one another, but that the lowest mean was significantly lower than the next lowest mean, at the .01 level. Consequently, a "cluster-straggler" conclusion is accepted. Two facts stand out in Table 3, the second more interesting than the first:

1. Mean set judgments are slightly higher

TABLE 1

MEAN JUDGMENTS OF SETS WITHIN CONDITIONS[a]

Heterogeneity of set components	
Minimum	3.43
Low	3.34
High	3.39
Maximum	3.04
Set mean scale value	
Low	2.02
High	4.58
Sex of photographs	
Male	3.34
Female	3.26
Number of judges per *J*	
One	3.26
Two	3.34
Sex of judges	
Male	3.35
Female	3.26
Order of presentation	
M-F	3.26
F-M	3.35
Set size	
Two	3.26
Three	3.38
Skewness	
No skew	3.38
Skew	3.26

[a] Based on all data, with exceptions, specified in Footnote 6 in which comparisons were based on data from two series only.

TABLE 2

SIGNIFICANCE TESTS OF MAIN EFFECTS[a]

Source of Variation	MS	*df*	*F*	*p*
Heterogeneity	5.88	3	10.45	<.001
Set mean scale value	1055.80	1	1876.40	<.001
Sex of photographs	4.59	1	8.16	.005
Number of judges per *J*	9.15	1	1.40	>.20
Sex of judges in *J*	7.83	1	1.20	>.20
Order of presentation	3.77	1	—	—
Set size	4.19	1	—	—
Skewness	4.13	1	—	—

[a] Based on all data, with exceptions, specified in Footnote 6 in which comparisons were based on data from two series only. Error Term II was used to test heterogeneity, set mean scale value, and sex of photographic subjects, while the other main effects were tested against Error Term I. See Footnote 4.

TABLE 3

COMPARISONS WITH RESPECT TO MEAN SCALE VALUE AND MEAN SET JUDGMENTS

	Series I (Level of MSV)		Series II (Level of MSV)		Series III (Level of MSV)		Overall (Level of MSV)	
	Low	High	Low	High	Low	High	Low	High
MSV	2.14	4.15	2.14	4.15	2.16	4.16	2.15	4.15
MSJ	2.10	4.42	2.04	4.71	1.91	4.61	2.02	4.58

than the mean of the scale values of the set components.

2. The sets of low MSV were judged below their MSV, while the sets of high MSV were judged above their MSV. The drop of the low sets was 0.13 and the rise of the high sets was 0.43, so that the gap between the two increased from 2.00 to 2.56, or about 25%.

The overall difference between MSV and mean set judgment was significant beyond the .001 level, as indicated by a *t* test. Additional *t* tests indicated that the drop in the judgments of low sets, the rise in the judgments of high sets, and the increase in the gap between conditions were all significant beyond the .001 level. It is not likely that these effects could be due to a biased selection of set components from the full scale, for the grand mean for the scale values of all 400 photographs was 2.95, which is not far from the mean MSV of 3.15 for all sets. Furthermore, the direction of this difference would favor a downward shift, contrary to the results obtained. Slightly more than half of the total number of component scale values on the full scale were between 2.15 and 4.15, the levels for the low MSV and the high MSV conditions, respectively, so the increased gap between conditions cannot be accounted for by the possibility that too small an initial difference in MSV levels was selected.

Since judgments of sets are systematically different from the mean of judgments of set components, we conclude that the condition of additivity, as Johnson (1955, pp. 306–308) calls it, which has also been referred to as "the linear hypothesis" by Rimoldi (1956), does not hold for stimuli of the type used. Johnson and Rimoldi refer specifically to the situation described by Equation 8 with $K = 0$, but because of the pronounced enhancement of contrast found, we may conclude that Equation 8 does not hold, no matter the value of K, for enhancement of contrast implies a slope greater than unity.

The remaining significant overall main effect is that for sex of photographic subjects. Sets of males were rated slightly higher than sets of females. Further results, discussed below, indicate that the interaction of sex of photographic subjects with sex of *J* is more pronounced than the main effect.

Analysis of variance. The data were subjected to extensive treatment by analysis of variance (cf. Willis, 1957). Briefly, the analyses consisted of (*a*) an analysis of Series I and II taken together, (*b*) an analysis of Series II and III taken together, and (*c*) an analysis of each of the series separately. The purpose of the first of these analyses was to test the significance of the set size factor and its interactions with certain other variables. The second analysis was performed in order to make similar tests about the skewness factor and its interactions. A primary advantage of the series-by-series analyses is that they make it possible to take into account the consistency of results running through three independent groups of *J*s, an important consideration. In fact, it seems reasonable to give as much or more weight to consistency as to level of significance per se. Because of the large number of *F* ratios computed (63 per series), very conservative standards were adopted for

accepting a finding as significant. The null hypothesis was rejected only for those sources of variation which were significant at or beyond the .05 level for *all three* series.[7]

The analyses of variance gave additional support for the significance of the main effects for heterogeneity and mean set value, but neither the main effect due to set size, nor that due to skewness, approached significance. Two interaction effects were found to be highly significant, however. First, the Series I–II analysis revealed that the difference in mean set judgments between the low and high MSV conditions is larger for 3-sets than for 2-sets, as shown in Table 4. This effect is significant beyond the .001 level. Not surprisingly, the effects of stimulus pooling were greater in the condition in which the larger number of stimulus components were pooled. It should be noted that the range of component scale values, both within high and low MSV conditions and overall, were very nearly the same for all three series, thus eliminating the possibility that differences in ranges could account for the obtained results.

Second, the series-by-series analyses showed that both male and female *J*s tended to be more lenient towards their own sex (Table 5). This pattern of differences held for all three series, reaching the .01 level of significance in Series I and III, and the .05 level in Series II. The differential is a little more than twice as large for male *J*s as it is for female *J*s.

Effects of number of judges per J on variance of judgments. In order to investigate the hypothesis that judgments made by partners exhibit less variance than those made by individual judges, mean variances were computed for the two *J* sizes. The mean variance for single judges was 37.97, while that for partners was 33.97. The difference, although in the expected direction, is not significant ($t = 1.20$, $df = 118$, $p > .20$). Pooling of responses did not produce a significant degree of convergence of judgments. Thus, even "the only consistent finding of significance for social psychology," cited by Johnson (1955, p. 471) in his discussion of group judgment, is not invariably found.

[7] For those sources of variation which were significant at or beyond the .05 level for *any two* of the three series, the null hypothesis was *tentatively* rejected. See the original thesis (Willis, 1957).

TABLE 4

SUBGROUP MEANS FOR INTERACTION BETWEEN MEAN SCALE VALUE AND SET SIZE[a]

Set Size	Mean Scale Value	
	Low	High
2	2.10	4.42
3	2.04	4.71

[a] Based on Series I and II.

TABLE 5

SUBGROUP MEANS FOR SEX INTERACTION BETWEEN SEX OF PHOTOGRAPHIC SUBJECT AND SEX OF JUDGES[a]

Sex of *J*s	Sex of Photographs	
	Male	Female
Male	3.48	3.21
Female	3.20	3.32

[a] Based on all data.

DISCUSSION

The most interesting results, from the point of view of social stereotyping processes, are (*a*) the highly significant enhancement of contrast, which resulted in an increased scale separation in the set-judging condition, and (*b*) the more pronounced nature of this effect with the sets having the larger number of components. These findings strongly suggest that, quite independently of learning and memory effects, there is a purely judgmental process that contributes to enhancement of contrast.

The enhancement of contrast observed here, however, is of a different kind from that found by Campbell (1956). The judges in his experiment made judgments about group members, not about groups as such. He found that the amount of overlap between two groups was considerably underestimated, a result which he had predicted from considerations of stimulus generalization. Neither his prediction nor his result involved an increased separation between group means, such as was found in

the present experiment. In fact, his analysis calls for a *decrease* in the difference in group means due to intergroup stimulus generalization, a prediction which was very roughly confirmed by his data. Thus, although intergroup stimulus generalization may have played some role in the present experiment, its effects, if any, were more than neutralized by effects operating in the opposite direction on the differences in group means.

What underlying mechanism, then, may account for this absolute exaggeration of group differences? One possibility is that it is the result of efforts on the part of *J*s to reduce the cognitive strain resulting from ambiguity by sharpening intergroup differences. It can be assumed that *J*s had more difficulty in forming definite impressions when judging sets than when judging individual components, and that this difficulty, or cognitive strain, was greater for 3-sets than for 2-sets. The intergroup sharpening, along with relative intragroup leveling, served to reduce the impression of ambiguity and the concomitant cognitive strain. We may hypothesize that a need to reduce cognitive strain may also be one of the important motivations in the formation of social stereotypes, as well as in other instances of sharpening, leveling, and enhancement of contrast occurring in social perception.

A second possible mechanism, not mutually exclusive with the one above, is that *J*s were making judgments in terms of estimated probabilities, so that increasing distances in either direction from some central point on the response continuum corresponded to increasing degrees of unusualness. Then, in the set-judging condition, component probabilities were componded in some way to yield a lower probability for the set. This relatively low set probability would evoke a response assigning the set to a position at relatively great distance from the central point on the response scale. To take an analogy from everyday experience, it would be judged somewhat unusual to see a woman six feet tall, but it would be judged more unusual to see two or three such women together. Unfortunately, the data at hand do not allow a choice between the two interpretations.

Whatever the psychological mechanism or mechanisms may be, it is easy to show what is necessary mathematically to remove the restriction of unit slope from Equations 7 and 8, and so allow either for the possibility of sharpening (slope > unity) or leveling (slope < unity) of group differences. By assuming that a different rate-of-change constant applies for each group size, we replace Gulliksen's single differential equation (Equation 1) by a set of equations. For groups of Size 1 and 2, we write

$$dv_1/dx_1 = k_1 \qquad [9]$$

$$dv_2/dx_2 = k_2 \qquad [10]$$

where subscripts identify group size, and v and x are, as before, subjective value and objective stimulus measure, respectively. Since x_2 equals the sum of the two x_1, Equations 3–5 become

$$v_i = k_1(x_i - x_0) \qquad [11]$$

$$v_j = k_1(x_j - x_0) \qquad [12]$$

$$v_{ij} = k_2(x_i + x_j - x_0) \qquad [13]$$

By algebraic operations similar to those used to arrive at Equations 7 and 8, the following expressions are reached.

$$s_{ij} = K_1(s_i + s_j) + K_2 \qquad [14]$$

$$s_{ij} = K_1\left(\frac{s_i + s_j}{2}\right) + K_2' \qquad [15]$$

where $K_1 = k_2/k_1$, $K_2 = k_2x_0 - (2k_2/k_1 - 1)C$, and $K_2' = \frac{1}{2}K_2$. Equation 14 applies when judges are requested to respond in terms of the *sum* of the component scale values ("How much would you like to have this portable phonograph plus this pen and pencil set?"), while Equation 15 applies when judges are requested to respond to the *mean* of the component scale values, as in the present experiment. When $K_1 > 1$, group differences are exaggerated, while when $K_1 < 1$, group differences are underestimated. For the data of Series I $K_1 = 1.15$, and $K_2' = -0.37$. For three components per composite stimulus, Equation 15 may be generalized to read

$$s_{ijk} = K_1\left(\frac{s_i + s_j + s_k}{3}\right) + K_2' \quad [15]$$

where $K_1 = k_3/k_1$, $K_2 = 2k_3x_0 - (3k^2/k_1 - 1)C$, and $K_2' = K_2/3$. For the data of Series II and III taken together, $K_1 = 1.34$, and $K_2' = -0.90$.

Another finding which may have implications for social stereotyping is the slight, but significant, drop in mean scale values produced by extreme heterogeneity of component scale values within sets. Because the effect is absent with moderate degrees of heterogeneity, predictions made on the basis of set means alone will be nearly as accurate as predictions taking into account individual component scale values; but because the effect appears with extreme component heterogeneity, knowledge of the distribution variance may, in extreme cases, contribute to predictive power. To see how this might apply to social stereotyping, consider the example—entirely hypothetical—that it is known that Jews do not differ from other groups on mean introversion-extraversion, but are more widely dispersed about this mean, with higher proportions both of loud, aggressive individuals and of quiet, retiring ones. Stimulus pooling effects could conceivably contribute to a heavier weighting at one end of the continuum, and in this way facilitate the formation of a group reputation, even though the group in question does not differ in the mean from other groups.

The greater leniency of *J*s towards the own-sex subseries is likely due largely to the fact that judges were rating the individuals portrayed as potential dates or mates, and consequently adopted more critical standards. A second possibility is that, when judging members of one's own sex, a process of group identification produced a tendency to judge less critically. Identification might also have caused judges to assign themselves a position on the scale, not necessarily consciously. Then, the more leniently a given judge judges his own-sex subseries, the higher a rating he implicitly assigns himself.

In our culture, attractiveness is accepted by both sexes as being more important in the case of young women than it is for young men. This suggests that the date-rating set would be more pronounced for males judging females, but that group identification would be more salient for females judging females. Because the difference in mean set judgments is more than twice as large for male *J*s as it is for female *J*s, it seems probable that the first process is the more important one.

Turning to the negative finding of principal interest, the failure to find a significantly smaller variance of judgments in the "partners" condition, the most probable explanation is the lack of any but arbitrary, ambiguous scale units. This situation allowed partners, as compared to individual judges, to shift to a different equivalence ratio between the stimulus scale and the response scale. Partners no doubt felt as much of an obligation as did individual judges to use more than a few of the response categories, and because of this they expanded the scale just about the required amount to compensate for the effect of response-pooling. Presumably, if the stimuli had been of a sort that more clearly defined scale units (if scale units had been inches or dollars, for example), and if the relationship of responses to stimuli had been more clearly indicated (as with anchors), the reduction in variance would have been more pronounced. Experimentation to test this interpretation is currently in progress.

SUMMARY

The effects of stimulus pooling on judgments of personal attractiveness were investigated in an orthogonal factorial design, using as stimulus components identification type photographs which had been previously scaled on attractiveness. Composite stimuli were constructed from these components by mounting sets of two or three photographs on cards. These composite stimuli were varied systematically along several dimensions. Judgments of sets were then obtained under conditions similar to those under which components had been scaled. The most clear-cut and firmly demonstrated positive results were:

1. Mean set judgments (MSJs) were not equal to mean scale values (MSVs) of set components—the additivity condition did not hold. Rather, for sets of high MSV, $\overline{\text{MSJ}} > \overline{\text{MSV}}$, while for sets of low MSV, $\overline{\text{MSJ}} < \overline{\text{MSV}}$.

2. This enhancement of contrast was significantly more pronounced for sets of three components than for sets of two.

3. Extreme heterogeneity of component scale values slightly lowered MSJs; moderate heterogeneity had no significant effect.

4. Judges were more lenient in judging the attractiveness of own-sex photographs than other-sex photographs.

Negative results worthy of note were:

1. Judges working in pairs produced MSJs which were not significantly different, either in mean or variance, from those made by judges working alone.

2. The main effect due to set size was not significant.

3. The main effect due to the amount of skewness of component scale values about set MSV was not significant.

The implications of the findings for (*a*) social perception, especially social stereotyping, and (*b*) group judgment were considered.

REFERENCES

ALLPORT, G. W. *The nature of prejudice.* Cambridge, Mass: Addison-Wesley, 1954.

BARTLETT, F. C. *Remembering.* Cambridge, England: Cambridge Univer. Press, 1932.

CAMPBELL, D. T. Enhancement of contrast as composite habit. *J. abnorm. soc. Psychol.*, 1956, **53**, 350–355.

CLARKE, R. B., & CAMPBELL, D. T. A demonstration of bias in estimates of Negro ability. *J. abnorm. soc. Psychol.*, 1955, **51**, 585–588.

DUNCAN, D. B. Multiple range and multiple *F* tests. *Biometrics*, 1955, **11**, 1–42.

GULLIKSEN, H. Measurement of subjective values. *Psychometrika*, 1956, **21**, 229–244.

JOHNSON, D. M. *The psychology of thought and judgment.* New York: Harper, 1955.

KEPHART, W. M. Negro visibility. *Amer. sociol. Rev.*, 1954, **19**, 462–467.

RIMOLDI, H. J. A. Prediction of scale values for combined stimuli. *Brit. J. statist. Psychol.*, 1956, **9**, 29–40.

WILLIS, R. H. Stimulus pooling and response pooling in social judgments. Unpublished doctoral thesis, Univer. of Wisconsin, 1957.

(Received April 13, 1959)

Journal of Abnormal and Social Psychology
Vol. 60, No. 3, 1960

ACHIEVEMENT MOTIVATION IN FANTASY AND BEHAVIOR[1]

DONALD M. BROVERMAN, EDWARD J. JORDAN, JR.,[2] AND LESLIE PHILLIPS
Worcester State Hospital and Institute of Human Development, Clark University

EVIDENCE that motives can influence the content of perception has led to attempts to measure motivation by means of projective fantasy. McClelland, a leader in this movement, has presented a technique of assessing such motives as need Achievement, need Affiliation, and need Power from responses to modified versions of the TAT. Experimental work (McClelland, Atkinson, Clark, & Lowell, 1953; Atkinson, 1958) has indicated positive relationships between these fantasy measures and indices of the same motives derived from overt behavior. Such results have been taken as evidence for what might be termed a general-expression model of motivation, i.e., the view that motives are concurrently reflected in both fantasy and overt behavior.

Recently, however, the validity of this formulation has been questioned. Lazarus, Baker, Broverman, and Mayer (1957) found that achievement fantasy decreased with increasing age and education. Application of a general-expression model to this result implies a parallel reduction in the behavioral expression of achievement motivation with increasing age and education. This implication appeared implausible to Lazarus et al. Hence, these authors proposed that fantasy might serve as an alternative or substitute channel for the expression of achievement motivation when this motive is blocked from behavioral expression in real life. The higher achievement fantasy in younger, less well educated *S*s was attributed to their inability to express achievement needs in behavior, while the older individuals were seen as more able to express such needs, with a concomitant reduction of fantasy about achievement.

This alternative-channels model leads to the prediction of an inverse relationship between achievement fantasy and behavioral attempts to achieve. It thus stands opposed to the general-expression model of McClelland et al. (1953) which predicts a positive relationship between these two indices of achievement motivation.

The present study is concerned with an experimental resolution of these conflicting models. In attempting such a resolution, it seemed to us that the kind of behavioral index of motivation to be used was of critical consequence. Most previous experimentation has used for this purpose performances on laboratory experimental tasks, but in the light of a recent discussion of "task relevance" by Vogel, Baker, and Lazarus (1958) we felt that the assessment of achievement motivation in behavior by such means was of questionable validity. Laboratory tasks such as solving anagrams, solving arithmetic problems, or crossing out letters on a page would appear to have little relevance to an individual's principal interests in life, and might well yield data more influenced by such extraneous factors as the individual's proneness to believe laboratory instructions or his need to please the examiner. Furthermore, this technique subjects all *S*s to the same task, disregarding individual differences in specific areas of interest. Even if some *S*s were highly motivated by a laboratory task of arithmetic problems, for example, many others probably would not be. Finally, many experiments have intentionally used tasks that are relatively simple to perform, so that intelligence would not confound presumed effects of achievement motivation in performance. However, evidence (Lewin, Dembo, Festinger, & Sears, 1944) suggests

[1] This investigation was supported by the Dementia Praecox Research Project, Worcester State Hospital and a research grant (M-896) from the National Institute of Mental Health, Public Health Service.

Thanks are due Melvin Feffer, Worcester State Hospital, and Morton Wiener, Clark University, for their many constructive criticisms. Stanley Kaden and Jerome Podell served as judges for the scoring of the Worcester Scale of Social Attainment as used in this study. Role-taking-task material was gathered and scored by V. Gourevitch. We should also like to thank Thaddeus Clapp through whose generous cooperation subjects were made available for use in this study.

D. M. Broverman participated in this research as a United States Public Health Service Postdoctoral Fellow in Clinical Psychology.

[2] Now at Veterans Administration Hospital, Durham, North Carolina.

that striving declines with relatively easy tasks, paralleling a decrease in level of aspiration.

The preceding considerations impelled us to measure the behavioral expression of achievement motivation by means of life history data, which appear much more likely to reflect behaviors toward goals that are important to the individual. Life history data are unique to each individual, obviating the possibility that an irrelevant task is being imposed on *S*. Furthermore, their use allows the individual to report on his attempts to achieve goals at a level of difficulty suitable to his achievement motivation.

The present study investigates the relationship between achievement fantasy and behavioral indices of achievement motivation taken from the area of employment, a central concern for our *S*s, who were adult males. A positive relationship between achievement motivation in fantasy and behavior would be taken as supporting the general expression viewpoint on this motive, while a negative relationship would support the alternative-channels formulation. In either event, the use of a behavioral index drawn from life rather than laboratory behavior was seen as lending greater weight to any evidence that might be obtained.

METHOD

Subjects

Thirty-seven members of a local church were used in this study. All *S*s were paid for their participation, but their pay went to the support of their parish. The sample was characterized by a wide range with respect to socioeconomic variables. Education varied from 5 to 19 years, with the median at 13 years. Annual income extended from $3,000 to $10,000, with a median of $5,000. Job level, as defined by the Dictionary of Occupational Titles (DOT), ranged from managerial and professional to semiskilled, with a roughly even distribution of *S*s in all job categories. Age ranged from 24 to 50 years, with a median of 36.

Procedures

*S*s were administered the Worcester Scale of Social Attainment (WSSA) in a group. In a later session individual inquiry into responses was conducted. The WSSA is a broad assessment of the degree to which an individual meets certain standards of accomplishment in society, and is discussed more fully by Phillips and Cowitz (1953). Two sections of this scale were selected for use in this study: Striving for Advancement and Job Level.

Striving for Advancement (WSSA, Item 7) taps the extent to which an individual actively seeks to better his employment position. Scoring on this item is based on answers to the following check list.

1. I would like to go to school to get a better job.
   (*a*) Do you have a particular school in mind?
   (*b*) Name of school?
2. I have visited plants looking for a better job.
3. I am interested in another job if it is offered.
4. I read the help-wanted ads every day.
5. I have asked my friends to look for jobs for me.
6. I am going, or have gone to school lately for job improvement.
7. Sometimes I think about looking for something better.
8. I have written letters to employers.
9. I am keeping my eyes open.
10. I have tried to get interviews.
11. I would like to have a better job.
12. I would like to learn a trade.
13. I expect to work for myself.
14. I would like to work for myself.

Responses to these questions permit classification of *S*s into following four categories:

1. Desire and intense activity, e.g., letters written about jobs; interviews applied for; contacts for better position actively sought.
2. Desire and tentative activity, e.g., is interested in another job if offered; does not seek a job; does not write although may look through ads; has vague long range goals.
3. Desire but no activity, e.g., says he would like a change but does nothing.
4. No desire and no activity.

*S*s were placed in one of these four categories by two independent judges, whose judgments correlated .98 (Pearson rank order correlation).

Behavioral manifestation of the achievement motive is typically understood as a kind of striving measure, or a measure of how much effort is involved in trying to achieve; it does not necessarily reflect actual achievement. At the same time, however, it seems possible that actual achievement may influence the amount of achievement fantasy produced by an individual. For this reason, *S*'s occupational level was used as a control for actual achievement in the area in which his behavioral expression of achievement motivation was examined. This control measure was taken from the Job Level item of the WSSA (Item 1). This item assigns scores to each level of occupation, as defined by the DOT. Higher scores are assigned to higher job levels.

The fantasy measure of achievement motivation was taken from the Role-Taking Task (RTT) devised by Feffer (1959). This is a projective test, based on Shneidman's Make-a-Picture-Story (MAPS). Three background settings were used: a factory scene, a living room scene, and one other chosen by *S* from the MAPS materials. A variety of figures (men, women, and children) were provided for *S* to place in these scenes. *S*s then told a story, using at least three figures, for each situation. After completion of the three stories, *S*s were presented with each situation again and instructed to retell their stories from the viewpoint of each character in the scene.

RTT stories are scored for motivational content by

several criteria. Four of these were used in the present study, since they were McClelland's need Achievement categories of Achievement Imagery (AI), Stated Need for Achievement (nAch), Anticipatory Goal States (GA+, GA−), and Affective States (G+, G−). In a sample of 120 TAT stories presented and scored by McClelland et al., this portion of the need Achievement score correlated .95 with total need Achievement. Hence, it seems reasonable to assume that scores based on these four categories of need Achievement are highly representative of total need Achievement, as scored by McClelland et al. (1953). Each *S*'s achievement fantasy score is the sum of themes in his RTT stories which satisfy the criteria of any of the above four classes.

RESULTS

Achievement fantasy scores were analyzed in a two-by-two factorial analysis of variance. Independent variables used were: Striving for Advancement (WSSA, Item 7) and Job Level (WSSA, Item 1). *Ss* in the two higher striving categories were compared with those in the two lower categories. Job Level was divided at the median into high and low groups. The requirement of proportionality between the four cells (Lindquist, 1953) necessitated the random elimination of three *Ss* from the appropriate cells. The resultant *N* of this analysis is 34. Correlation between Striving and Job Level was negligible ($r = .11$).

Table 1 reveals that Striving is reliably associated with fantasy-expression of achievement motivation ($F = 4.35$, p less than .05). Table 2 indicates that the direction of this relationship is inverse: high strivers show low achievement fantasy, and low strivers show high achievement fantasy.

Table 1 further indicates that Job Level and fantasy-expression of need Achievement are not significantly related.[3]

TABLE 1
ANALYSIS OF VARIANCE OF ACHIEVEMENT FANTASY
($N = 34$)

Source	*df*	*MS*	*F*
Striving	1	79.53	4.36*
Job Level	1	4.63	
Interaction	1	4.99	
Within Error	30	18.22	
Total	33	SS = 635.88	

* F significant beyond the .05 level, two-tailed test.

TABLE 2
MEAN ACHIEVEMENT-FANTASY SCORES OF HIGH AND LOW STRIVING GROUPS

High Striving ($n = 17$)	4.11
Low Striving ($n = 17$)	6.47

DISCUSSION

This study demonstrates an inverse relationship between behavioral striving and achievement fantasy,[4] and thus supports an alternative-channels view of the expression of achievement motivation. Interpretation of these results must take account of the numerous studies which support the general-expression model by demonstrating positive relationships between achievement fantasy and striving. The alternative-channels view can no more account for these instances of positive relationships than the general-expression model can account for the present results. We are inclined to attempt a resolution of this conflict of results via an extension of the alternative-channels model.

We would approach this problem by suggesting that persons unable to express their achievement needs in life-situations will be impelled to seek alternative modes of need expression, not only in fantasy, as suggested by Lazarus et al. (1957), but also in behavioral striving when tasks simpler and less ambiguous than life problems are encountered.

The experimental tasks used to assess behavioral striving typically are safer and/or better structured than the situations in which most life goals are embedded. Also, because they are made easy, intelligence presumably plays little, if any, part in achievement on them. Therefore, persons who are blocked in expressing achievement motivation in real

[3] Two other possible indices of achievement motivation, educational level and annual income, also failed to predict achievement fantasy. While it is impossible to draw general conclusions from these findings, it appears that actual achievement was not a predictor of achievement fantasy in the present sample.

[4] Vogel, Baker, and Lazarus (1958) also report an inverse relationship between achievement fantasy and a behavioral index of achievement motivation (performance on the McKinney Reporting Task). However, this relationship was found only with fantasy indices obtained after a failure-stress condition; no relationship was found with fantasy obtained under neutral conditions. Hence, the relevance of this finding to the theoretical issue discussed in this paper is unclear.

life might be expected to show a great deal of striving on such tasks. On the other hand, *Ss* who are able to express their achievement needs in life should have no need to exert themselves on tasks so distant from real-life concerns. Such persons might be expected to show little striving in the laboratory. It then follows from the present finding, i.e., a negative relationship between life striving and achievement fantasy, that low laboratory strivers would be the low achievement fantasiers, while high laboratory strivers would be the high achievement fantasiers. Thus, a reversal of striving patterns from life to the laboratory may account for the positive relationships previously observed between striving and achievement fantasy. Such a reversal occurs, according to this argument, because performance of laboratory tasks, like fantasy, may serve as a substitute expression of blocked achievement motivation.

The extended alternative-channels model, then, seems preferable to the general expression model for several reasons. It can account for both positive and negative relationships between striving and achievement fantasy and it is supported by evidence drawn from life history data. Finally, this model is generally consonant with many current developments in psychological theory: the psychoanalytic emphasis upon thought and fantasy as consequences of delayed action (Rapaport, 1951); evidence presented by the Lewinian school for substitute activity and cosatiation (Lewin, 1935); and demonstrations of the vicarious relationship of fantasy to action by developmentally oriented studies (Krus, Werner, & Wapner, 1953; Hurwitz, 1954; Kruger, 1954; Misch, 1954).

Nevertheless, it is clear that any attempt at a definitive statement about the relationship between achievement fantasy and striving would be premature at present. The extension of the alternative-channels model advanced in the present paper has not yet been empirically tested. Furthermore, additional factors such as cognitive controls (Klein, 1954; Lazarus et al., 1957) and environmental conditions (Birney, 1959), may influence the kind of relationships that are obtained.

One such factor, particularly pertinent to this study, is variation in fantasy-assessment technique. The RTT, on which our results are based, differs from the usual achievement-oriented TAT by allowing *S* some choice in determining the test stimuli and by requiring the retelling of stories from several viewpoints. While it is difficult to envision how these differences might reverse the direction of relationship between striving and fantasied achievement, this possibility cannot be ruled out, and our findings must be qualified in this regard.

In any event, the present study casts doubt on the unqualified assumption that achievement motivation is generally expressed and supports the conceptualization of fantasy and behavior as alternative channels for the expression of this motive.

SUMMARY

This study investigates the relationship of achievement fantasy to behavioral indices of achievement motivation. Conflicting views of this relationship have been presented. McClelland et al. (1953) have assumed that fantasy is a direct index of a generally expressed motive state and have supported their position with findings of positive relationships between fantasy and behavioral indices of achievement motivation. Lazarus et al. (1957) on the other hand, have conceptualized fantasy as an alternative, substitute channel for the expression of achievement motivation when behavioral channels are blocked. This latter view implies an inverse relationship between fantasy and behavioral expression of motivation.

The present experiment is an attempt to determine which of these views is applicable when behavioral indices of achievement motivation are derived from life history data.

Thirty-seven adult males, representing a wide range of age and socioeconomic variables, were administered the Worcester Scale of Social Attainment (WSSA) and a Role-Taking Task (RTT). Two behavioral indices of achievement motivation were taken from the WSSA: Job Level, a measure of actual achievement, and Striving for Advancement, a measure of the degree to which an individual is actively trying to raise his level of occupational achievement. A fantasy measure of achievement motivation was derived from the RTT.

An inverse relationship was found between the striving and fantasy measures of achievement motivation. No significant relationship was found between actual achievement and either achievement fantasy or striving.

These results were interpreted as supporting the alternative-channels model of Lazarus et al. (1957). Previous research in this area was criticized, and an attempt was made to extend the alternative-channels model to cover the reported instances of positive relationships between achievement fantasy and behavioral expression of motivation.

REFERENCES

ATKINSON, J. W. *Motives in fantasy, action, and society.* Princeton, N. J.: Van Nostrand, 1958.

BIRNEY, R. C. The reliability of the achievement motive. *J. abnorm. soc. Psychol.*, 1959, **58**, 266.

FEFFER, M. The cognitive implications of role-taking behavior. *J. Pers.*, 1959, **27**, 152–168.

HURWITZ, I. A. A developmental study of the relationships between motor activity and perceptual processes as measured by the Rorschach test. Unpublished doctoral dissertation, Clark Univer. 1954.

KLEIN, G. S. The personal world through perception. In R. R. Blake, & G. V. Ramsey (Eds.), *Perception, an approach to personality.* New York: Ronald, 1951.

KRUGER, A. Direct and substitute modes of tension-reduction in terms of developmental level: An experimental analysis by means of the Rorschach test. Unpublished doctoral dissertation, Clark Univer., 1954.

KRUS, D., WERNER, H., & WAPNER, S. Studies in vicariousness: Motor activity and perceived movement. *Amer. J. Psychol.*, 1953, **66**, 603–608.

LAZARUS, R. S., BAKER, R. W., BROVERMAN, D. M., & MAYER, J. Personality and psychological stress. *J. Pers.*, 1957, **25**, 559–577.

LEWIN, K. Substitute activity and substitute value. In *A dynamic theory of personality.* New York: McGraw-Hill, 1935.

LEWIN, K., DEMBO, T., FESTINGER, L., & SEARS, P. S. Level of aspiration. In J. McV. Hunt (Ed.), *Personality and the behavior disorders.* Vol. 1. New York: Ronald Press, 1944. Chap. 10.

LINDQUIST, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.

MCCLELLAND, D. C., ATKINSON, J. W., CLARK, R. A., & LOWELL, E. L. *The achievement motive.* New York: Appleton-Century-Crofts, 1953.

MISCH, R. The relationship of motoric inhibition to developmental level and ideational functioning: An analysis by means of the Rorschach test. Unpublished doctoral dissertation, Clark Univer. 1954.

PHILLIPS, L., & COWITZ, B. Social attainment and reactions to stress *J. Pers.*, 1953, **22**, 270–283.

RAPAPORT, D. Toward a theory of thinking. In D. Rapaport (Ed.), *The organization and pathology of thought.* New York: Columbia Univer. Press, 1951.

VOGEL, W., BAKER, R. W., & LAZARUS, R. S. The role of motivation in psychological stress. *J. abnorm. soc. Psychol.*, 1958, **56**, 105–112.

(Received April 30, 1959)

Journal of Abnormal and Social Psychology
Vol. 60, No. 3, 1960

ATTITUDINAL CONSEQUENCES OF COMMITMENT TO UNPLEASANT BEHAVIOR[1]

JACK W. BREHM
Duke University

THE theory of cognitive dissonance (Festinger, 1957) has led to empirical investigation of the psychological consequences of inconsistencies between attitudes and behavior. It has been shown, for example, that a person induced to argue against his private opinion will tend to show change in his private opinion toward the induced position (Cohen, Brehm, & Fleming, 1958; Festinger & Carlsmith, 1959). It has also been shown that simply engaging in a behavior inconsistent with one's attitudes results in attitude change toward consistency with the behavior (Aronson & Mills, 1959; Mills, 1958). Other derivations concerning the consequences of inconsistency between attitude and behavior have yet to be tested. This paper reports a test of a relatively "nonobvious" derivation.

According to the theory, when a person is induced to engage in some behavior that he ordinarily would avoid, he experiences dissonance. He then tries to reduce the dissonance, and, in the present case, one way he may do so is by changing his attitude toward the behavior (or some aspect of it) in such a way as to reduce the reasons for avoiding it. Thus, an individual induced to perform a *disliked* behavior should tend, among other things, to *increase his liking* for that behavior.

The person's tendency to change his attitude is conceived to be a direct function of the magnitude of dissonance experienced. Given his dislike of a behavior and with the inducing force held constant, the magnitude of dissonance increases as the amount of committed behavior increases. And given the fact that he has committed himself to some disliked behavior, the less support that further information to which he is exposed gives to that behavior, the greater the magnitude of dissonance.

Specifically, then, the following derivations may be made.

1. The greater the amount of disliked behavior to which a person commits himself, the greater the magnitude of dissonance and consequent increase in liking for the behavior or the behavioral object. This hypothesis has been tested by Aronson and Mills (1959), who showed that the more severe a person's initiation in joining a group, the greater his liking for the group members and their activities.

2. Given commitment to disliked behavior, the less supportive is additional information, the greater the magnitude of dissonance and consequent increase in liking. This effect is expected to increase as a function of the amount of committed behavior. The present study was designed to test this derivation.

The theory also asserts that a person experiencing dissonance tries to reduce it and to avoid increasing it. Thus, the extent to which information either for or against engaging in the behavior is believed should be a function of the amount of committed behavior. This consideration gives rise to another derivation tested by this study.

3. Supporting information is accepted and nonsupporting information is rejected by a person in proportion to the amount of disliked behavior to which he is committed.

METHOD

The general procedure has been described in a previous report (Brehm, 1959). Briefly, it consisted of (*a*) a pre-experimental measure to ascertain a vegetable disliked by each of the subjects (*Ss*), (*b*) an individual experimental session in which each *S* was induced to commit himself to eating some of the disliked vegetable, (*c*) exposure of the *S*, after he was committed, to information in support of or against eating the disliked vegetable, and (*d*) measurement of relevant affect and opinions before and after the experimental inductions.

The *pre-experimental questionnaire* was given to eighth-grade students in their English classes.[2] It consisted of a response scale for each of 34 vegetables. *S* was to indicate how much he liked each vegetable by an X on any one of 61 dots running along a straight

[1] This project was conducted as part of the Yale Communications Research Program. It was supported by a grant to Carl I. Hovland from Bell Telephone Laboratories.

[2] For their help in conducting this study, the author wishes to thank Louisa Willard, Psychologist for the New Haven Schools, Frank Carr, Principal of Troupe Junior High School, Teacher Connolly, and the eighth grade teachers, especially Teacher Himmelfarb.

line and identified at intervals of 10 dots by the following labels: like very much; like a lot; like a little; neither; dislike a little; dislike a lot; dislike very much. *Ss*' responses were then scored from +3.0 for "like very much" to −3.0 for "dislike very much."

Experimental procedure. Each *S* was assigned, during school hours, to go to the Homemaking room of the school where he was met by the experimenter (*E*). It was explained that the project was "consumer research" about various people's preferences for different kinds of vegetables. *E* said that he wanted to ask the *S* to answer some questions about a vegetable picked at random (though it really had been selected as one quite disliked, at least −2.0 on the pre-experimental questionnaire). After answering some preliminary questions, *S* was given the main questionnaire to fill out. The questions, to be described in detail, were designed to measure liking for the vegetable and opinions about its food value.

On completion of the main questionnaire, *S* was told that his answers would be very helpful but that a person sometimes gave different answers if he first tasted the vegetable. It was further explained that *S* need not do anything further, since he might not like the idea of eating some of the vegetable, but that he would be given a reward if he chose to eat the vegetable and answer further questions about it. At this point the experimental manipulation of amount of committed eating was introduced.

Commitment to little eating (Low Eating condition) was accomplished by telling *S* that to obtain the reward he must eat a single small dish of the vegetable immediately. Commitment to greater eating (High Eating condition) was accomplished by telling *S* that to obtain the reward he must eat a small dish of the vegetable immediately and also return three or four times within a few weeks, each time to eat more. All experimental *Ss* were offered the same reward, a gift certificate good for their own choice of two 45 rpm or 78 rpm records, or two tickets for movies of their own choice. *Ss* were assigned at random to the two commitment conditions. It should be noted that the overall rate of refusal to eat was only 5% and did not differ between conditions.

If *S* agreed to continue, he was given a small dish of the vegetable, heated if appropriate, and allowed a few minutes in which to eat it. When finished, the *S* was given a "research report that he might be interested in reading." It was ostensibly from the laboratories of the testing organization and purported to give the food value of the vegetable. Actually, each report was designed to give information in support of (Support condition) or against (No Support condition) eating the vegetable. The following report, for example, was designed to *support* the eating of spinach.

> Vegetables have two important values in a person's diet—they furnish most of the minerals and most of the vitamins needed for strong bodies and active minds.
>
> Our tests of spinach show that this vegetable is particularly strong on vitamins. An average size portion will furnish 95% of the daily quantity of vitamins needed for building new body tissues and mending injured tissues and bones. Spinach is probably the best single vegetable to supply vitamins.

Other reports in the Support condition were identical except for the name of the vegetable. Reports in the No Support condition were identical except that they said "weak" instead of "strong," "only 5%" instead of "95%," and "worst" instead of "best."[3] The type of report given to each *S* was determined by chance.

When *S* had finished reading the report he was asked to fill out again the main experimental questionnaire (identical to the "before" questionnaire). In making the request, *E* mentioned that some people change one way, some another, and some don't change at all. He also reaffirmed the anonymity of the *S*'s answers.

Upon completion of the questionnaire, *S* was allowed to choose his reward and was asked not to talk about what he had done when he returned to class. Shortly after the completion of the project, *E* met with all the *Ss* in a group and explained the deceptions and true purpose of the project.

Control procedure. In order to estimate any direct effects of the communications without a commitment to eat, a classroom experiment was performed. For this test 8th graders from a different school had to be used.[4] The second school is not far from the first and draws its students from a population with similar socioeconomic status. The same preliminary questionnaire was given, though this time by the *E* and only one week prior to the experimental session. A set of materials was prepared for each *S*, and consisted of the main questionnaire under which was a sealed booklet made up of either a Support or No Support communication for a vegetable disliked by the *S*, and a second copy of the main questionnaire. These were passed out to the students in class and *E* explained he was trying to obtain more detailed information by asking each person several questions about one particular vegetable whose name he would find on his questionnaire. *Ss* were asked first to tear off their names, which were on separate top sheets. The questionnaires, thus, looked anonymous although they were coded on the back. *E* then asked *Ss* to fill out the first questionnaire carefully but not to unseal the booklet until told to do so. When the first questionnaires were completed and passed in, *E* explained that while this information would be good, he had also found that people sometimes gave different answers if they thought carefully about the vegetable first, trying to imagine as vividly as possible just how it tastes, smells, and feels in the mouth. So, he said, he was also asking them to do this and then to answer some further questions they would find on the inside of the booklet. Also, on the inside of the booklet they would find a laboratory report about their vegetable in which he thought they would be interested, and they should read that before answering the questionnaire. After all

[3] About half of the *Ss* received similar communications concerning mineral instead of vitamin content. Since this type of communication failed to affect liking either directly or in conjunction with commitment, the results for these subjects are omitted from this report.

[4] The author wishes to thank Louisa Willard, Principal Corrigan, Principal of Truman School, and Teachers Walker, Smernoff, and Kelly, at Truman School, for their help in conducting this part of the study.

final questionnaires had been completed and turned in, the true purpose of the experiment was explained.

The main experimental questionnaire. This questionnaire included one question identical to that used on the prequestionnaire, which read, "Considering everything about it, (taste, smell, appearance, texture) to what extent do you like or dislike this vegetable?" It was answered on a scale like that used on the prequestionnaire. A second question pertained to the food value of the vegetable, and read, "How high would you guess the general vitamin content of this vegetable compared to others?" The answering scale was similar to that used for the previous questions except that the identifying labels were: very high; fairly high; slightly above average; just average; slightly below average; fairly low; very low.

RESULTS

Commitment and change in belief. While the content of the communications was direct and forceful and the source was highly authoritative, *S*s might still manage to believe them differentially. According to dissonance theory, the greater the magnitude of dissonance being experienced, the greater the seeking of information that can reduce dissonance and the avoidance of information that might, increase it. Thus, persons in the High Eating condition, more than those in the Low Eating or Control conditions, should tend to raise, or resist lowering, their estimate of vitamin content.

Mean changes in estimated vitamin content are given in Table 1. The means for all three conditions show a decrease in the No Support and an increase in the Support condition, the difference being statistically reliable for each. It is clear that the manipulation of Support and No Support was successful. Mean changes in the High Eating condition, however, show less reduction in estimated vitamin content from the No Support communication, significantly different from the Low Eating condition at the .07 level, and from the Control condition at beyond the .01 level.[5] Similarly, the increase in estimate from the Support communication is greater in the High Eating condition than in either Low Eating or Control conditions, though not significantly so. If the effects of the Support and No Support communications are averaged, the High Eating condition is significantly different from the Low at the .04 level, and from the Control at beyond the .01 level. It is apparent, then, that *S*s tended to believe supporting information and disbelieve nonsupporting information in proportion to the amount of eating to which they were committed.

TABLE 1

MEAN INITIAL SCORES AND CHANGES IN ESTIMATED VITAMIN CONTENT

Commitment	Communication			
	No Support		Support	
	Initial	Change (*N*)	Initial	Change (*N*)
Control	.58[a]	−3.02 (11)	.84	0.92 (10)
Low Eating	.45	−2.35 (8)	1.76	0.77 (7)
High Eating	1.23	−0.34 (7)	.46	1.66 (10)

[a] Comparisons of the initial means do not yield significant *t*s, indicating that the initial scores are somewhat unstable with *N*s this small.

Effect of support on liking. The communications were given in a class room situation to estimate their effect on liking for the vegetable. It was assumed that any such effects obtained in the Control condition would be at least as large in magnitude for the individually tested *S*s is the main experimental conditions.

Mean changes in liking (from the "before" to the "after" main experimental questionnaire) were computed for the Control conditions. These changes, shown in the first row of Table 2, indicate that the Support communication produced increased liking, the No Support decreased liking, the difference between the conditions being significant at the .05 level.

Effect of commitment and support on liking. Since the Control condition showed communication *support* to increase liking while the dissonance hypothesis holds that *nonsupport* increases liking, the two effects are working against each other. Thus, in the Low Eating condition, where the magnitude of dissonance is low, we might expect the support manipulation to have a direct effect on liking, similar to that in the Control condition. But in the High Eating condition, where the magnitude of dissonance is greater, the effect of attempts to reduce dissonance should reduce or even reverse the direct effect of the support manipulation on liking.

Mean liking change scores for each experimental condition are presented, along with the Control data, in Table 2. The direct effect of

[5] All *p* values are from two-tailed *t* tests.

TABLE 2
MEAN INITIAL LIKING AND CHANGES FOR EACH SITUATION

Commitment	Communication			
	No Support		Support	
	Initial	Change	Initial	Change
Control	−.52[a]	−.43[b]	−1.56	.46
Low Eating	−.99	.11	−.66	1.97
High Eating	−.03	1.33	−.64	.48

[a] None of the below comparisons yields a significant t for initial scores.

[b] p values for various comparisons of change scores:
Control (No Support-Support) = .05
Low Eating (No Support-Support) = .06
High Eating (No Support-Support) = .09
No Support (High Eating-Control) = .02
No Support (High Eating-Low Eating) = *NS*
Support (Low Eating-Control) = <.01
Support (Low Eating-High Eating) = <.01
Low Eating (No Support-Support)–High Eating (No Support-Support) = .02
Control (No Support-Support)–High Eating (No Support-Support) = .03

the support manipulation occurs in the Low Eating condition: increase in liking is greater for the Support than for the No Support communication (significant at the .06 level). And as expected from the dissonance hypothesis, the difference between Support and No Support conditions with the High Eating manipulation is reversed. While the difference is significant at only the .09 level, it is significantly different from that for the Low Eating condition at the .02 level, and from the Control condition at the .03 level. It is clear, therefore, that the direct effect of supporting information on increased liking tends to be reversed to the extent that there is behavioral commitment.

It will also be noted in Table 2 that commitment has no over-all effect on liking. The Low Eating conditions produced as much increased liking on the average as did the High Eating. These results therefore fail to provide adequate support for the hypothesis that increased liking is a simple function of amount of eating committed.

DISCUSSION

This experiment has shown that a person induced to perform a disliked behavior tends to increase his liking for the behavior as a joint function of the amount of the behavior to which he is committed and the presence of further supporting or nonsupporting information. The major results are in support of Festinger's theory of cognitive dissonance. When a person is induced to perform a disliked behavior, additional information can decrease or increase the magnitude of dissonance by supplying information that is respectively for or against engaging in the behavior. He attempts to reduce the dissonance, in proportion to its magnitude, by increasing his liking for the behavior. Thus, he increases his liking for a behavior to which he is committed to the extent that he is exposed to information against engaging in the behavior.

While the predicted interaction was obtained, the manipulation of amount of committed eating failed to produce an over-all effect on liking. Average change in liking in the Low Eating condition was as great as in the High. One possible reason for this unexpected result is that *S*s in the High Eating condition, who were expecting to return again, were more restrained in their general reaction to the situation. Another possibility is that the Eating manipulation was not sufficiently strong to provide an overall effect; only in conjunction with the No Support communication did it produce dissonance. The latter interpretation, however, would not explain why with the Support communication the Low Eating condition produced greater increased liking than did the High Eating condition. A third possible interpretation is that the pressure to change liking is a function of the discrepancy between behavior and information. That is, where behavior is minimal and informational support high, or behavior is great and support low, there is pressure to increase liking. When, however, both behavior and support are great, the pressure to increase liking is low. Thus, it is suggested that dissonance may be created by exposure to supporting information, and may be reduced interchangeably by commitment to behavior or by increased liking. This implication of the theory is not developed in Festinger's statement (1957).

The experiment, it should be noted, was concerned with only one possible channel of dissonance reduction. It assumed that attitude change toward the induced behavior would be a major mode of reduction under the present conditions, though other modes were available

to subjects and may have been used to some degree. While this assumption was supported by the positive findings of the study, it might not prove to be a good assumption in other settings. As yet, dissonance theory does not completely specify the factors controlling the choice of channel through which dissonance is reduced. Making the theory operational therefore requires taking such extratheoretical factors into account. There is here a clear need for theoretical refinement and further research. Meanwhile, the results of the present experiment can be generalized to other situations only with due caution.

SUMMARY

This study tested a "nonobvious" hypothesis derived from Festinger's theory of cognitive dissonance: given the fact that a person is committed to an unpleasant behavior, he tends to increase his liking for that behavior more if he is exposed to information against engaging in it than if he is exposed to information favorable to engaging in it. This tendency is anticipated to be in proportion to the amount of such behavior to which he is committed.

The hypothesis was tested by inducing junior high school students to commit themselves to eating different amounts of disliked vegetables. Fictitious research reports to which they were exposed said their vegetable was either high or low in food value. Estimates of food value and liking for the vegetable were measured before and after the experimental inductions.

It was found that the communications about food value tended to be believed or disbelieved as a function of the amount of eating to which the person was committed. In addition, the main experimental hypothesis, as stated above, was supported.

REFERENCES

ARONSON, E., & MILLS, J. The effect of severity of initiation on liking for a group. *J. abnorm. soc. Psychol.*, 1959, **59**, 177–181.

BREHM, J. W. Increasing cognitive dissonance by a *fait accompli*. *J. abnorm. soc. Psychol.*, 1959, **58**, 379–382.

COHEN, A. R., BREHM, J. W., & FLEMING, W. H. Attitude change and justification for compliance. *J. abnorm. soc. Psychol.*, 1958, **56**, 276–278.

FESTINGER, L. *A theory of cognitive dissonance*. Evanston: Row, Peterson, 1957.

FESTINGER, L., & CARLSMITH, J. M. Cognitive consequences of forced compliance. *J. abnorm. soc. Psychol.*, 1959, **58**, 203–210.

MILLS, J. Changes in moral attitudes following temptation. *J. Pers.*, 1958, **26**, 517–531.

(Received July 6, 1959)

Journal of Abnormal and Social Psychology
Vol. 60, No. 3, 1960

THE FREQUENCY OF SELF-REFERENCE STATEMENTS AS A FUNCTION OF GENERALIZED REINFORCEMENT[1]

J. STACY ADAMS AND BROWNING HOFFMAN
Stanford University

RECENT experiments have demonstrated applications of operant conditioning in the realm of verbal behavior (e.g., Hildum & Brown, 1956; Verplanck, 1955; see Krasner, 1958, for review). Stimuli demonstrated to be reinforcing have been verbalizations of assent, agreement and approval, nods, smiles, paraphrases and impersonal events such as buzzer and bell tones and colored lights. Attention, although hypothesized by Skinner to be a generalized reinforcer (Skinner, 1953, 1957) has not been specifically investigated.

The purposes of this study were (*a*) to determine whether a combination of attention and assent constituted a reinforcing event, (*b*) to explore the effects of prolonged extinction on the occurrence of self-reference statements and other behavior such as emotional reactions and escape behavior, and (*c*) to determine what are the implications of the use of common events (attention-assent) in free, unstructured interview situations. In most previous studies extinction has not been carried out at length. In Verplanck's investigation, for example, extinction was carried out for only 10 minutes, a period too brief to uncover any effects other than those upon the class of operants considered (Verplanck, 1957).

METHOD

General Experimental Plan

The experiment was conducted under the guise of being a survey of student opinion with two groups of 31 *S*s, one experimental group and one control group. The survey interview was divided into four consecutive periods, the first three each lasting 10 min., while the last varied in length up to a maximum of 20 min. Each period was further subdivided into 2-min. intervals during each of which *S*'s self-reference statements and other statements were counted. After the introduction of the "survey" and after *S* understood what was wanted of him, the first 10-min. period began and the operant level of statements containing self-references was determined, no reinforcement being given in either group.

During the second 10-min. period *E* reinforced all statements containing self-references made by experimental *S*s by following their occurrence by a combination of attention and assent. A precise description of reinforcing events is given below. Control *S*s received no reinforcement in Period II, nor in any other period.

Extinction was conducted during the third 10-min. period and during the subsequent 20-min. period. As will be seen, some *S*s stopped responding entirely during the course of extinction. Extinction consisted of withholding from experimental *S*s the reinforcement previously given in Period II.

If at any time during the course of the experiment *S*'s over-all rate of responding fell to zero, nondirective probes such as "What else can you think of?", "Can you think of anything else?", or "Do you have any other ideas?" were used. These were infrequently necessary except during extinction and then they were often unsuccessful in maintaining continued responding.

Subjects

The experimental *S*s were 31 male and female, graduate and undergraduate students at Stanford University, who were "interviewed" in the spring of 1958. Control *S*s were drawn from the same population but were "interviewed" in the spring of 1959. It was ascertained at the conclusion of the experiment, when the real purpose of the "interview" was explained, that all *S*s were naive to the extent that they had no training in conditioning and had never before participated in an experiment on operant conditioning.

Experimenters

The *E*s were advanced undergraduate and graduate students in two of the senior author's courses on language and verbal behavior. Seven of these collected the experimental group data; two collected the control group data. One *E* collected data from both groups. In addition to having some theoretical sophistication in verbal operant conditioning, each *E* had a minimum of four complete practice sessions followed by a detailed critique period. The *E* who collected data from both groups had, in addition, one year's intensive practice in verbal conditioning in connection with a number of other studies. He acquired considerable sensitivity in identifying statements and self-reference statements and imparted this skill to the other *E* who collected control data and for whose training he was responsible. As a consequence, to which reference will be made below, rates of response show an elevation in the control group data as compared to the experimental group data.

In view of the fact that there is a difference in the sensitivity of the experimental and control group *E*s, it

[1] The assistance of Beverly A. Brown, George E. Dunn, Ben V. Hole, Frank B. Krasne, Lillian Lauer, Jack Lyle, and Aline Spivock is gratefully acknowledged.

would have been desirable to obtain interexperimenter reliability data. Unfortunately, any useful reliability data would have had to be based on the reinterview of the same *S*s by different *E*s and this, of course, would have had an effect on the observed responses. Our feeling is that *E*s of equivalent training would have yielded acceptably high reliabilities. Since the seven *E*s who collected the experimental group data had comparable training and the two who collected control group data had equivalent training, the inter-*E* reliabilities within the two groups are probably acceptable. And this is really the important point, for we are especially concerned with change or lack of change *within* each group.

Procedure

The experiment was conducted at the convenience of *S*s in their living quarters or other quiet place at a time when only *E* and *S* were present and when *S* had one full hour available. When the experiment proper began, *E* was seated in front of *S* and recorded the frequency of self-reference statements and all statements at the end of each 2-min. period.

In his introduction to *S*, *E* told him that he was doing a "survey" for the psychology department, that the purpose was to find out students' "frank and honest personal opinions and ideas about conformity and education in America," and that *S* could talk for about an hour or more on the subject. *S* was further told that *E* would not enter into the interview except to take notes. *S* was then reminded that the interview would be most fruitful if a complete, detailed and exhaustive expression of his ideas and opinions were obtained.

The topic "conformity and education in America" was selected for several reasons. First, it was a topic students frequently discussed and about which they held quite definite opinions. Thus, it was felt that there would be some motivation to express views on it. Secondly, the subject was broad enough to permit sustained talking. Thirdly, it was a topic which neither imposed an upper nor a lower limit on the use of statements of self-reference.

Except during the 10-min. reinforcement period for the experimental *S*s, *E* kept his entire attention on his notebook and was silent. He did not lift his head, shift his line of regard to *S*, or make any verbal responses, unless *S* stopped talking. When the latter occurred *E* used a general, nondirective probe in order to try to maintain *S*'s output.

Reinforcing Stimulus Events

A single class of stimulus events was hypothesized to have reinforcing properties. The event consisted of *E*'s (*a*) lifting his head, (*b*) shifting his line of regard from his notebook to *S*, and (*c*) murmuring "mm-hmm" simultaneously. After a reinforcement (which took about 1 sec.) *E* returned to his original position and composure until *S* emitted another statement containing a self-reference. As previously indicated, it is only among experimental *S*s in Period II that reinforcement of self-references was used. Among control *S*s, *E* was silent throughout the experiment and kept his attention on his notebook.

The Class of Operant Responses Conditioned

The verbal operants selected for conditioning were all statements containing one or more references to *S* personally. Thus, statements having "I," "me," "my," "mine," "myself" and like words were considered self-reference statements, except in those instances when the pronoun had someone else than *S* as a referent (e.g., "my roommate thinks . . . "). Statements containing the first person plural pronoun, "we," were also counted as self-reference statements if they pertained to an activity in which *S* himself had participated, or a belief he shared.

Statements were defined as units of "thought," containing a complete idea. Statements, therefore, were complete sentences as defined in standard dictionaries, although in speech they might not be clearly punctuated as a written sentence. Such a definition is not entirely satisfactory, for the sufficient and necessary criteria are not enumerated, and there is, therefore, room for subjective variability on the part of *E*s. *E*s, however, become more proficient with experience and there is a tendency for them to identify more self-reference statements as their proficiency increases.

RESULTS

Effects of Reinforcement

The basic index of responding used in the analysis is the ratio of self-reference statements to total statements made (called SR ratio hereafter). The mean SR ratios for each period and the mean SR ratio over-all periods are plotted in Fig. 1 for the experimental and control *S*s.

The first point to be noted is that the mean over-all SR ratio in the experimental group is .305, whereas it is .408 in the control group. The difference is significant beyond the .001 level and represents increased proficiency by one *E* in identifying statements and self-references as he gained experience during a period of approximately one year, as previously related.

As Fig. 1 shows, the size of the SR ratio increases during reinforcement and drops during extinction in the experimental group, as predicted under the hypothesis that attention-assent constitutes a reinforcing state of affairs. To determine the significance of the observed change in SR ratios two analyses were performed. First, Cochran's *Q* test (Siegel, 1956) was applied to determine whether the values of SR deviated significantly from the mean over-all SR value of .305. This required a determination of whether each individual ratio deviated above or below that value. The *Q* value

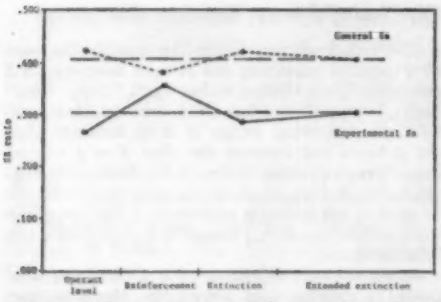

FIG. 1. SR ratio by period for experimental and control subjects.

obtained was 9.38, which is significant beyond the .05 level. Among the controls the SR values during each period did not differ significantly from the over all mean SR value of .408 ($Q = 2.68$).

The second analysis, Friedman's nonparametric analysis of variance (Siegel, 1956), reveals that the treatments contributed significantly to the variance among experimental *Ss* ($\chi_r^2 = 16.7$) but not among control *Ss* ($\chi_r^2 = 5.15$). Further analysis of the data for the experimentals revealed that the reinforcement treatment is the major contributor. Wilcoxon's signed-ranks analysis for matched pairs (Siegel, 1956) indicated that the SR ratio during reinforcement among experimental *Ss* was significantly greater (*a*) than the SR ratio during the operant level period ($z = 2.5$), (*b*) than the SR ratio during the first extinction period ($z = 2.9$), and (*c*) than the SR ratio during the second extinction period ($z = 3.1$).

The rate of emitting self-reference statements and all statements is shown in Fig. 2. Among the experimental *Ss* the rate of self-reference statements increases significantly from 1.42 to 1.90 self-reference statements per minute during the reinforcement period ($z = 2.98$, by Wilcoxon's test), while the rate of emission of all statements remains virtually constant. Among the control *Ss* both the rate of self-references and the rate of all statements drops significantly during the corresponding period ($z = 3.20$ and $z = 3.13$, for self-reference and all statements, respectively).

Effects of Extinction

The SR ratio among experimental *Ss* decreases significantly following reinforcement, during the first extinction period ($z = 2.9$). In the second extinction period, lasting up to 20 min., there is no further decrease in the SR ratio (Fig. 1). However, the rate of self-reference statements decreases continuously during extinction, as may be seen in Fig. 2. By Wilcoxon's test, the rate of self-reference statements during the first extinction period is significantly smaller than during reinforcement ($z = 3.4$) and is significantly smaller during the second extinction period than during the first ($z = 3.08$). The rate of making all statements is not significantly smaller during the first extinction period than during the reinforcement period ($z = 1.27$). However, the rate is significantly smaller during the second extinction period than during the reinforcement period, ($z = 4.3$). Furthermore, the rate during the second extinction period is smaller than during the first extinction period ($z = 3.96$). Therefore, although the SR ratio during the second extinction period is not smaller than during the first extinction, the rate of emitting self-references and total statements decreases during successive extinction periods.

The rates of self-reference and all statements for control *Ss* do not vary significantly after the first 10-min. period of the "interview." The extinction effects noted among experimental *Ss* are in no way evident among the controls. Hence, the observed experimental extinction effects are attributable to withdrawal of generalized reinforcement.

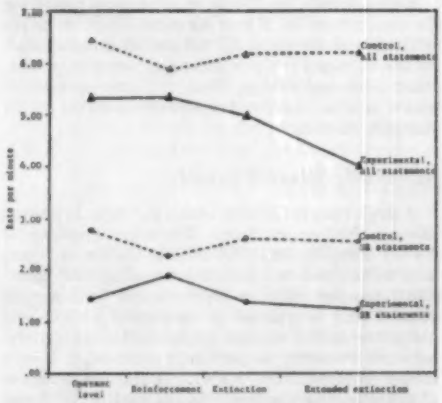

FIG. 2. Rate of emitting self-reference and all statements per minute for experimental and control subjects.

Many experimental *S*s displayed general disturbance, emotionality, and, in some instances, downright anger and hostility toward *E* during extinction. This was especially noticable in the late stages of extinction. In her report on her "interviews," one of the experimenters wrote: "... many *S*s found the interview to be a stress situation (during extinction)." This finding is in accord with reports by Verplanck (1955). There was further evidence that some *S*s "left the field"—i.e., refused to continue speaking, despite prompts—because the discontinuance of reinforcement was aversive to them. It is as though nonreinforcement constituted *punishment* for them. However, it is not possible to say that all *S*s whose verbal behavior stopped before the end of the second extinction period did so because of aversion. In only one *S* was there evidence of emotionality among the controls, though a number of them stopped responding before the end of the 50-min. experimental session, despite the use of probes, apparently having run out of comments on the subject of conformity.

In Fig. 3 the number of *S*s who continued to respond during each 2-min. interval of the extinction period is plotted. At the beginning of the extinction period all experimental *S*s were responding; during the corresponding period twenty-eight of the thirty-one controls were responding. The difference is not significant by Fisher's exact test ($P = .12$, one-tailed). At the end of the extinction period, however, nine experimental *S*s and sixteen control *S*s were responding. The difference is significant beyond the .05 level ($\chi^2 = 3.28$, one-tailed).

Evidence of Awareness

Upon questioning at the end of the experiment not one experimental or control *S* reported that the "interview" situation was other than was related to him. Some experimental *S*s reported "awareness" of change in *E*'s behavior during the reinforcement and extinction periods, this change being one of "interest" and "attention." None, however, were able to report on what these changes were contingent—i.e., emitting self-reference statements. We may conclude, therefore, that the observed experimental effects were not a function of knowledge or awareness of the experimental manipulations.

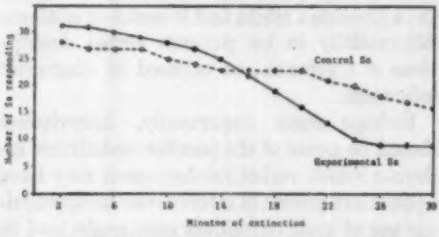

FIG. 3. Number of experimental and control subjects continuing to respond at various two-minute intervals of extinction.

DISCUSSION

The evidence is unequivocal that the frequency of emitting self-references can be controlled by making attention-assent contingent upon their occurrence. When attention and assent follow, self-references increase (relative to total output); when attention and assent are withdrawn, the relative frequency of self-references decreases. The same is true for the rate of emitting self-reference statements. These effects are not observable among the controls, whose self-reference statements were not reinforced at any time. Further evidence of the effects of the experimental manipulations comes from the fact that *all* verbal behavior is extinguished in significantly more experimental than control *S*s.

Four explanations of the latter over-all extinction phenomenon suggest themselves. First, it is plausible that *S*s stop responding because they have exhausted all that they can say on the subject of conformity and education in America. Though this is undoubtedly true in the case of some *S*s, since a number of control *S*s stopped responding, such an explanation does not account for the fact that the experimental *S*s are more likely than the control *S*s to stop responding before the termination of the experiment. Nor does it account for the experimental *S*s' greater emotionality during extinction. Second, it is possible that statements without self-references were extinguished along with those containing self-references because of generalization. Both types of statements are quite similar in that both are verbal responses and it is therefore possible that extinction of self-reference statements generalizes to other statements. However, in the present instance the explanation appears improbable

because, if there were stimulus generalization, it should have been evident during reinforcement. But there was no evidence for stimulus generalization during reinforcement, for only the frequency of self-reference statements and rate of emitting them increased, while total output and rate of total output remained constant.

A third possibility is that a minimum number of self-reference statements is necessary if talking on the selected topic is to be possible. There is no direct evidence in support of this explanation. However, the data in Fig. 1 indirectly suggest that it may have validity. After prolonged extinction the SR ratio among experimental *S*s drops only to the operant level in those *S*s who continue talking. Thus, approximately .270–.300 may be the *minimum* possible ratio if talking is to continue. If the proportion of self-references drops below this level (because of extinction) talking may become impossible or at least very difficult. Further support for this hypothesis comes from the fact that while the SR ratio remains constant during extinction, the *rate* of self-references and total output steadily decreases. Among the controls, on the other hand, both the SR ratio and the rate of self-reference statements and of total output remain constant during the corresponding periods.

A fourth possible explanation of the extinction of all verbal behavior is that withdrawal of reinforcement constitutes punishment and that this punishment has the effect of arresting all verbal behavior (and perhaps other behavior) in the presence of *E*. If punishment is defined as "an operation in which an aversive or conditioned aversive stimulus is made contingent upon a response" (Ferster & Skinner, 1957, p. 731), the explanation seems reasonable, for silence is probably used by the verbal community as a conditioned aversive stimulus. As Skinner points out, remaining silent when the occasion usually gives rise to speech constitutes punishment (Skinner, 1957). Thus, in the present experiment the *E*'s withholding of attention and assent, following reinforcement, may have had an effect very similar to the silence following the commission of a verbal *faux pas* in a social group. That some experimental *S*s who stopped responding may have perceived the situation in precisely this way is evidenced by their reporting that they felt the *E* had become uninterested in or even annoyed at what they were saying and by the fact that they frequently became visibly tense, angry, and hostile.

Of the four possible explanations suggested, the data rule out that relating to the generalization of extinction. Running out of comments to make may partially account for the *S*s stopping to respond altogether, for some *S*s in the control group, as well as in the experimental group, stopped before 50 min. elapsed. However, it could not account for the significantly greater proportions of experimental *S*s who stopped. This disparity may be accounted for by either the third and fourth explanations offered, or, perhaps, by both. The data do not permit a choice between the two, however.

With respect to interviewing of the free, unstructured type, the findings of this study have clear implications. Attention and assent—both common interviewer responses—can produce significant changes in the content of a respondent's verbalizations. In this study, the effect was to increase the frequency of self-reference statements. When one member of a dyad differentially reinforces specific verbal response classes of the other participant, he gains some control over the conversation. Such control has obvious advantages. Psychotherapists may use verbal reinforcement techniques to guide their patients toward significant insights; these procedures may be of considerable importance, for instance, when applied to self-reference, affect, or other statements where therapeutic circumstances preclude direct requests for such information from the patient. Similarly, survey interviewers can control the content of a respondent's verbalizations in a direction consonant with the purpose of an interview by use of differential reinforcement. Or, a physician might find it useful to reinforce differentially in his patients verbal descriptions of symptoms, as opposed to diagnostic inferences.

Perhaps more importantly, interviewers should be aware of the possible undesirable influence which verbal reinforcement may have upon the responses of interviewees. Inappropriate use of such techniques may easily lead to selective biases in the data or clinical evaluations based upon reflected idiosyncracies of

the interviewer which have been reinforced in a respondent or patient. Furthermore, the withdrawal of reinforcement, as in the present study, may result in emotionality that might have undesirable consequences, particularly in psychotherapeutic situations.

SUMMARY

An experiment was conducted to determine whether the use of a combination of attention and assent by an "interviewer" constituted a reinforcing event and to find out the effects of prolonged extinction. Following a 10-min. period during which the operant level of self-reference statements was measured, all self-reference statements by experimental *Ss* were reinforced by *E*'s showing attention-assent for a period of 10 min. The *Ss*' responses were then extinguished for a period of as long as 30 min. The same procedure was followed with control *Ss* with the exception that they received no reinforcement.

It was found that attention-assent constituted a reinforcing state of affairs since the relative frequency and rate of self-reference statements significantly increased during conditioning among the experimental *Ss*, while control *Ss* showed no change. Extinction of self-reference statements produced a significant decrease in their relative frequency and rate of occurrence among experimental *Ss*, but not among control *Ss*.

During extinction most experimental *Ss* showed emotional disturbance; in some cases anger and hostility. The verbal behavior of significantly more experimental than control *Ss* was totally extinguished before the end of the 30-min. extinction period. In some instances it was evident that *Ss* were leaving an aversive field. Possible explanations of the observed extinction phenomena were offered.

There was no evidence that any *S* was aware or had knowledge of the experimental manipulations.

REFERENCES

FERSTER, C. B., & SKINNER, B. F. *Schedules of reinforcement*. New York: Appleton-Century-Crofts, 1957.

HILDUM, D. C., & BROWN, R. W. Verbal reinforcement and interviewer bias. *J. abnorm. soc. Psychol.*, 1956, **53**, 108–111.

KRASNER, L. Studies of the conditioning of verbal behavior. *Psychol. Bull.*, 1958, **55**, 148–170.

SIEGEL, S. *Nonparametric statistics*. New York: McGraw-Hill, 1956.

SKINNER, B. F. *Science and human behavior*. New York: Macmillan, 1953.

SKINNER, B. F. *Verbal behavior*. New York: Appleton-Century-Crofts, 1957.

VERPLANCK, W. S. The control of the content of conversation: Reinforcement of statements of opinion. *J. abnorm. soc. Psychol.*, 1955, **51**, 668–676.

(Received July 1, 1959)

Journal of Abnormal and Social Psychology
Vol. 60, No. 3, 1960

INTERPERSONAL PERCEPTION AND COMMUNICATION IN DISCUSSION GROUPS OF VARIED ATTITUDINAL COMPOSITION[1]

IRWIN ALTMAN[2] AND ELLIOTT McGINNIES
University of Maryland

In a recent paper, the investigators reported a series of measures that distinguished between groups with homogeneous attitudes discussing a persuasive film on the topic of ethnic minorities (McGinnies & Altman, 1959). The present study applied these and additional measures to groups differing in attitudinal composition so as to determine the relationships existing between group structure and discussion behavior. Certain of Festinger's assumptions concerning forces to communicate in groups were also testable in the design of the experiment.

The structural properties of groups, and particularly the relative homogeneity of member orientations toward an issue, have been shown to influence perceptual and judgmental processes of the members (Sherif, 1932; Asch, 1952). In many of these studies, no systematic analysis of behavior was undertaken during the actual process of social interaction. We have attempted to follow influence pressures in different group structures throughout the process of member interactions, as well as to describe some relationships between group composition and interpersonal perception.

Recent studies (Festinger, 1950; Festinger & Thibaut, 1951; Festinger, Back, Schacter, Kelley, & Thibaut, 1950) have focused on the process of group functioning and have analyzed changes in member behavior as they influence and are influenced by others in the group. In general, these studies support the hypothesis of a positive relationship between degree of opinion discrepancy concerning an issue and the relative strength of pressures to communicate with those holding deviant opinions. In the majority of these studies, however, the relative balance of divergent opinions within groups has not been independently controlled. For example, some studies use four-man majorities facing one-man minorities, while others have much larger majorities. Since group composition has been shown to be an important determinant of member behavior, we controlled the degree of opinion discrepancy in groups by systematically varying majority-minority conditions.

Several additional differences between the present study and others dealing with intragroup pressures should be noted. Most studies in discussion settings have used situations where members have had to rely on others in the group for confirmation of their views. In the present study, certain members of the group found their views supported by the content of a persuasive film presented as a topic for discussion. In addition, earlier studies have defined opinion discrepancy as the degree of separation of members along an attitude continuum. Rather than manipulate the degree of distance between individuals in each group, we varied the number of individuals who represented extreme ends of an ethnocentrism continuum.

Briefly, the aims of the present study were as follows:

1. To explore the effects of various degrees of minority-majority conditions on member behavior during the course of group discussions.

2. To determine the relationships among attitudes, interpersonal perceptions, and forces to communicate within the different types of groups.

METHOD

Independent Variables

A slightly modified version of the California E scale was administered to 500 male students of the junior and senior classes at two high schools.[3] On the basis of

[1] This research was supported in part by a special grant from the National Institute of Mental Health, United States Public Health Service, Department of Health, Education, and Welfare, and is based upon portions of a doctoral dissertation submitted by the first author to the University of Maryland.

[2] Now with Human Sciences Research, Inc., Arlington, Virginia.

[3] We are grateful to the faculty and staff at Bladensburg and High Point Senior High Schools, Maryland, for their assistance in the conduct of this research. John Lesser, Robert Lana, and Richard Page assisted in the collection of data and management of the discussion sessions.

their scores, *Ss* were randomly assigned to one of five types of discussion groups, each of which consisted of six members. A total of 25 groups were formed, five in each of the following structures:

Low E homogeneous: All six members had low E scale scores.

Low E majority: Four *Ss* were low Es; two were high Es.

Balanced groups: There were equal numbers of high and low Es.

High E majority: Four members were high Es; two were low Es.

High E homogeneous: All six group members were high Es.

Each group met on a single occasion 3–4 weeks after the initial attitude screening. The first part of the meeting was devoted to viewing a film dealing with the problem of minority groups. This film, "The High Wall," takes an antiprejudice position in dealing with the conflict between teenagers of different ethnic backgrounds. Following the film, the group discussed the general problem of ethnic conflict. The discussion was led by a member of the research team who assumed a nondirective role and stimulated rather than guided the discussion. During the discussion, the investigators monitored recording equipment and coded the sequence and direction of participant responses. From both these "spottings" and recordings, we were able to compute a number of indices of discussion behavior.

At the conclusion of each discussion, group members completed a questionnaire dealing with attraction to the group, attraction to particular group members, importance of the discussion topic, and perceived consensus of opinion. In addition, *Ss* were readministered the E scale.

Dependent Variables

In order to assess the effect of various group structures on member behavior, three classes of dependent measures were used:

1. *Discussion Behavior.* Several discussion behavior indices previously developed (Gerard, 1953; Dickens, 1955; McGinnies & Altman, 1959) were employed. These measures required minimal inference by observers and were readily derived from the recorded discussions.

(*a*) *Spontaneity.* In discussion situations characterized by permissive leadership, opportunity is afforded group members to display varying degrees of initiative in their discussion behavior. The extent to which members act independently of leader stimulation is referred to as spontaneity. Operationally, a remark which occurred without leader probing was rated as spontaneous, whereas member responses which followed leader questions, probes, etc. were defined as nonspontaneous. A percentage index of spontaneity was obtained by computing the number of spontaneous remarks relative to the total number of responses.

(*b*) *Rate of response.* This general activity indicator is computed by determining the number of responses emitted by group members (or groups as a whole) per unit of time.

(*c*) *Spread of participation.* In discussions where there is no requirement that members participate, it is probable that inequalities in response output will occur. To compare the homogeneity of verbal output in the various structure conditions, a measure of spread of participation was employed (Dickens, 1955).

(*d*) *Direction of response.* Since we recorded the originator and the recipient of every discussion comment, it was possible to chart the direction of communication flow in the various groups. A modification of a measure developed by Gerard (1953) was used to represent response-direction scores.

2. *Attitude Change.* The E scale was administered several weeks prior to the film-discussion session and again immediately following the discussion. Although our major aim was to assess variations in discussion behavior and interpersonal perceptions as a function of group composition, we used this opportunity to relate discussion behavior to shifts in attitude.

3. *Questionnaire Data.* In addition to the E scale, a questionnaire consisting of measures of attraction to the group, sociometric preference, relative importance of the discussion topic, and accuracy of interpersonal perception, was administered at the conclusion of each discussion. Some of the more interesting and significant findings emerged from analysis of these data.

RESULTS

Discussion Behavior

(*a*) *Spontaneity.* This index, which refers to the degree of initiative of group members, was computed by dividing the number of spontaneous participations by the total number of participations during each of three equal time periods. Indices were computed for each group as a whole as well as for high and low Es separately. Since the meetings lasted approximately one-half hour, each index was based on about ten minutes of discussion.

A repeated-measures analysis of variance indicated significant changes in spontaneity over the course of the meetings for all groups, as well as for high and low Es considered separately. All groups exhibited more initiative as the meetings progressed (Fig. 1). Although no reliable differences in spontaneity appeared

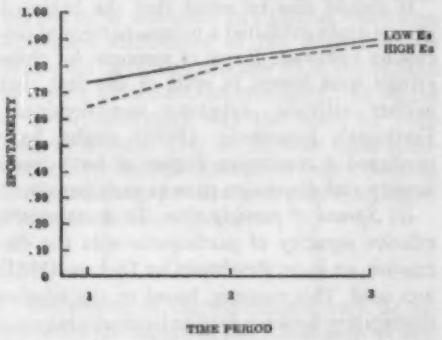

FIG. 1. Spontaneity indices over time periods.

among the various structure conditions, the low Es were generally more independent of the discussion leader than the high Es. These results are consistent with those obtained earlier by McGinnies and Altman (1959). In addition, there was a tendency for the balanced groups to exhibit lower spontaneity scores than any of the other groups, suggesting that structures balanced with respect to opinion heterogeneity lead to a reduced level of member initiative.

(*b*) *Rate of response.* A measure of discussion pace, or rate of output, was computed by dividing the number of participations in each of three time periods by the length of the time period and adjusting for the number of members in groups and subgroups. The index was computed for each total group, as well as separately for high and low Es.

Analysis of response rates showed no significant changes over the course of a meeting for either high or low Es. Moreover, no significant variations in discussion activity occurred for the different group structures. While statistical evidence is lacking, the data of Figure 2 suggest that all group members quickened the place of their discussion behavior as the meetings progressed. As shown in the figure, the low Es had generally higher rates of participation than the high Es ($t = 4.08$, $p < .01$). Thus, low Es tended to be both more active and less dependent on the leader than the high Es. It is interesting to note, however, that the rate of response difference stems largely from the heterogeneous structures. When there is absence of opinion divergence on the discussion topic, there is practically no difference in participation rates of high and low Es.

It should also be noted that the balanced groups again exhibited a unique pattern of discussion behavior. Rates of response for these groups were lowest in spite of the fact that neither attitude subgroup was dominant. Festinger's hypothesis (1950) might have predicted a maximum degree of both spontaneity and discussion pace in such groups.

(*c*) *Spread of participation.* To measure the relative equality of participations in the discussion, an index developed by Dickens (1955) was used. This measure, based on the relative discrepancy between each individual's response

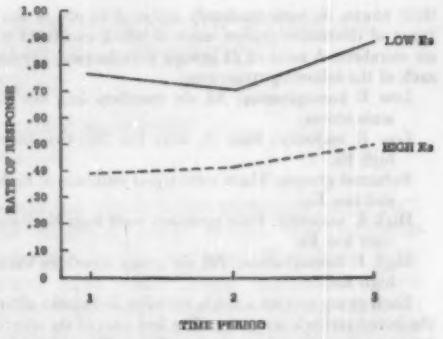

FIG. 2. Rate of response over time periods.

output and the group mean, yields an indication of the extent to which certain individuals dominated the discussions.

While Kruskal-Wallis analyses of variance showed no differences in relative spread of participation among the various group structures for either high or low Es, there were significant differences between the two attitude groups. Low Es tended toward equality in their individual contributions to discussions, while the high Es participation was dominated by a few individuals.

(*d*) *Direction of response.* The directional flow of communications between members was computed by dividing the number of responses directed toward a given subgroup by the product of the number of possible recipients, the number of members in the communicator subgroup, and the length of a time period. These *D* scores were computed for each of the possible communication channels in the heterogeneous structures: high E to low E; low E to high E; low E to low E; and high E to high E.

These scores were analyzed by Friedman's analysis of variance to ascertain temporal changes in communication flow and by the Kruskal-Wallis procedure to evaluate group structure and attitudinal differences in response direction.

When analyzed temporally, no differences in direction of communication flow occurred. High and low Es did not direct their remarks differentially over time. However, there were significant differences in communication flow for each of the four channels between the various structure conditions. The results show, first, that there was a paucity of opposition directed

communication in the balanced groups. When no clearcut majority existed, members tended to direct their remarks largely to those participants who held views similar to their own. When considered with the rate of response and spontaneity data, we arrive at the picture of balanced groups being relatively lethargic or apathetic in their discussions.

A second set of results indicate that low Es tended to increase communication toward both high and low Es as their number decreased, i.e. as they moved toward the minority position. Conversely, high Es tended to decrease opposition-directed remarks as they progressed toward the minority position. Thus, there is evidence for the presence of an interaction between attitude and group structure as determinants of influence processes.

Attitude Change

Since the E scale was administered to *S*s several weeks before discussions and again after each meeting, it was possible to compare changes in attitudes across the different group structures. Analysis of covariance of post-discussion scores indicated that neither high nor low Es changed differentially in ethnocentric attitudes as a function of group structure. However, there was a significant decrease in ethnocentrism for pooled *S*s ($p < .01$).

Questionnaire Responses

In addition to the E scale, participants completed a short questionnaire following the discussions. The questionnaire covered a number of general topics including (*a*) the extent to which members were attracted to their group, (*b*) the personal relevance of the discussion topic, (*c*) the attraction of members to one another, and (*d*) the perception of opinions held by other members.

(*a*) *Attraction to the group.* Although low Es were least attracted to the balanced groups, they did not exhibit differential attraction to the various group structures ($p > .05$). On the other hand, high Es did show significant attraction differences, with the lowest degree of positive feeling occurring in the balanced groups. Thus, both types of *S*s were least positive in attraction to their groups in those structures where neither attitude subgroup was in the majority. One other general finding, though not significant, was that low Es seemed to be more frequently attracted to their groups than the high Es.

(*b*) *Relevance of the discussion topic.* Although there was a slight tendency for low Es to place greater importance on the discussion topic, neither predisposition group responded differentially to this item across the various group structures.

(*c*) *Sociometric choice.* By having each *S* rank the other members with respect to personal attraction, we obtained another measure of the extent to which *S*s perceived one another's attitudes as these were revealed in the course of group discussion.

Although analysis of variance showed no difference between the choice patterns of high and low Es, there were significant differences in choices among the various structure conditions and a significant attitude × structure interaction. Examination of the data shows that the interaction results from the fact that the low Es were more apt to choose other low Es in both the minority and majority positions, while the high Es showed no appreciable differences across structures. In other words, the low Es reacted sociometrically to differences in attitudinal composition of the groups. The high Es apparently were imperceptive of such differences, at least so far as sociometric relationships were concerned.

(*d*) *Perceived consensus of opinion.* We obtained a measure of the accuracy with which member attitudes were perceived by having each *S* identify those participants whom they thought held opinions similar to and divergent from their own.

The results showed that low Es were generally more accurate in identifying the attitudes held by other members. In addition, there were significant differences in the degree of accuracy for both high and low Es in the various structure conditions. Interestingly, the source of these differences was the same for both attitude groups. That is, *S*s were most sensitive to attitudinal differences when judging from a minority position and were least sensitive in the balanced groups.

DISCUSSION

The results of the experiment may be examined with respect to our initial aims as well

as to certain hypotheses suggested earlier by Festinger (1950).

Festinger predicts, for example, that increased pressure to communicate to others in a group occurs with increases in the perceived opinion discrepancy among members. Our data enabled us to determine the extent to which the group members accurately perceived one another's ethnocentric bias as well as the direction, rate, and spontaneity of their discussion behavior. Since the low Es were more accurate than the high Es in identifying the attitudes held by others in their group, it follows that the low Es should have been more impelled to communicate than the high Es, who were less cognizant of intragroup discrepancies in attitude. Our results support this prediction, since the low Es displayed higher rates of response than the high Es. The data show that this difference in response rates between high and low Es derived mainly from groups in which the low Es were in a minority, and it was precisely these groups that the low Es were most sensitive to attitudinal differences among the group members. Of course, the generally greater discussion activity of the low Es could have been due in part to the fact that the communication supported their point of view.

An analysis of the low E response rate shows an increased tendency for these *S*s, when in the minority position, to interact with both low Es and high Es. It might have been predicted that the low Es, with film support of their views, would have directed the preponderance of their comments toward the "deviant" high Es. This, in fact, was the case, although they also increased the number of remarks directed toward those holding similar views. Thus, as their number decreased, and as they became more aware of the opinions held by the other members of the group, the low Es increased their rate of communication to both the opposition and to those who agreed with them. Apparently, communication behavior is not designed solely to influence those who disagree with one's own views, but also aims at providing and receiving support from those holding similar views.

We obtained support for Festinger's hypothesis that increased pressure to communicate with others occurs with increases in the cohesiveness of the group. Low Es, who had the highest rates of response, were more attracted to their groups than high Es. Attraction to the group for the low Es, however, was not related to group structure. In the case of the high Es, the greatest attraction to the group was reported by those in the majority position; and it was in these groups that they communicated more with the low Es than they did with each other. Attraction to the group, then, to extend Festinger's suggestion, is apparently an important predictor not only of the amount, but of direction of communication.

On the sociometric item, the high Es divided their choices about equally between low Es and other high Es, regardless of group structure. The low Es, however, showed significantly greater preference for one another when they were in either a majority or a minority position. We suspect that this may be a reflection of the more active discussions held under these two conditions, which gave the participants a greater opportunity to assess one another's attitudes. In fact, it was in these groups that the perceived consensus judgments by low Es were most accurate. Having assessed fairly accurately the ethnocentric dispositions of other members, the low Es preferred those individuals whom they perceived as more like themselves. The high Es were not as perceptive of differences in group structure and, therefore, could not be expected to assign their sociometric choices on other than a 50–50 basis.

Finally, we may note the interesting behavior displayed by the balanced groups, which were composed equally of high and low Es. Under these conditions, spontaneity, rate of response, and number of opposition-directed comments were generally lower than in any of the other structures. It may have been predicted that maximum opinion discrepancy would be associated with greater discussion activity. Our findings also indicate that the members of the balanced groups were least attracted to their groups, were least accurate in judgments of each other's attitudes, and were less likely to feel compatible with those individuals holding opinions similar to their own. These results suggest that where opinion equilibrium exists in a group, discussion activity is suppressed, with a resultant loss in accuracy of interpersonal perception and group

cohesiveness. One is tempted, in view of the previous finding by McGinnies and Altman (1959) that middle E groups were less active in group discussion than either low Es or high Es, to speculate that something analogous to an algebraic summation occurs when equal numbers of low Es and high Es form a group. A balancing of attitudes seems to produce discussion behavior similar to that in neutral groups.

SUMMARY

The purpose of this study was to investigate the effects of various degrees of minority-majority conditions on member behavior during the course of group discussions and to ascertain some of the relationships between various aspects of interpersonal perceptions and discussion behavior.

On the basis of their scores on a modified version of the California E scale, high school students were randomly assigned to one of five discussion conditions representing various degrees of group composition. Two group structures had homogeneous membership with respect to ethnocentric attitudes, that is, all had either high or low scores on the California E scale. The remaining three group structures were heterogeneous with respect to E scale scores and contained either low Es in the majority, high Es in the majority, or equal numbers of each.

Each six-man group viewed and discussed a film dealing with the general problem of ethnic minorities and then completed a questionnaire measuring various aspects of interpersonal perceptions. Discussion behavior was measured by a series of indices coded during the course of group meetings.

With one exception, the results showed no clear-cut and simple group structure effects, but indicated an interaction between group composition and interpersonal perceptions as determinants of discussion behavior. The major exception occurred for groups composed of equal numbers of high and low Es. In these groups, spontaneity, rate of response, and number of opposition-directed communications were generally lower than in any of the other structures. Members of these groups also were least attracted to their groups, were least accurate in their perceptions of opinions held by other members, and were less likely to choose sociometrically others in the group holding views congruent to their own. These findings suggest that opinion "equilibrium" results in a suppression of discussion activity and concomitant pressures to influence members holding divergent points of view.

The results of the study indicated an intimate linkage between interpersonal perceptions and discussion behavior. Those members who were most accurate at identifying the opinions held by others in the group played a larger role in discussions than those who were not so accurate. This effect was most pronounced in the heterogeneous structures, where there were divergencies of opinion. In addition, there was a tendency for discussion pace and opposition-directed communication to be positively related to valuation of the group.

The findings lend some support, as well as some qualifications, to Festinger's conclusions regarding influence effects in small groups.

REFERENCES

ASCH, S. E. *Social psychology*. New York: Prentice-Hall, 1952.

DICKENS, M. A statistical formula to quantify the "spread-of-participation" in group discussion. *Speech Monogr.*, 1955, **1**, 28–31.

FESTINGER, L. Informal social communication. *Psychol. Rev.*, 1950, **57**, 271–292.

FESTINGER, L., BACK, K., SCHACTER, S., KELLEY, H., & THIBAUT, J. *Theory and experiment in social communication*. Ann Arbor, Mich.: Institute for Social Research, 1950.

FESTINGER, L., & THIBAUT, J. Interpersonal communication in small groups. *J. abnorm. soc. Psychol.*, 1951, **46**, 92–99.

GERARD, H. B. The effect of different dimensions of disagreement on the communication process in small groups. *Hum. Relat.*, 1954, **3**, 313–327.

MCGINNIES, E. M., & ALTMAN I. Discussion as a function of attitudes and content of a persuasive communication. *J. appl. Psychol.*, 1959, **43**, 53–59.

MITNICK, L. L., & MCGINNIES, E. M. Influencing ethnocentrism in small discussion groups through a film communication. *J. abnorm. soc. Psychol.*, 1958, **56**, 82–90.

SHERIF, M. A study of some social factors in perception. *Arch. Psychol.*, New York, 1935, **27**, No. 187.

(Received June 2, 1959)

Journal of Abnormal and Social Psychology
Vol. 60, No. 3, 1960

CONDITIONING AND GENERALIZATION OF GSR AS A FUNCTION OF AWARENESS[1]

B. B. CHATTERJEE AND CHARLES W. ERIKSEN
University of Illinois

In 1937, Diven reported an experiment purporting to demonstrate the conditioning of anxiety reactions in human subjects without their awareness of the conditioned stimulus. His procedure involved a list of stimulus words in which the word "Barn" occurred six times. The subjects (*Ss*) were required to chain associate to each of these stimulus words, and following their chain associations to the word Barn *Ss* were given a strong electrical shock. Diven reported that *Ss*, although unable to verbalize the contingent relationship between barn and shock, gave significantly greater GSRs to the word Barn than did a control group of *Ss* who had not been shocked. Further, he reported that the conditioned GSR showed semantic generalization to other rural words that had been contained in the stimulus list. Haggard (1943) in a somewhat similar experiment obtained comparable results.

For obvious reasons, these two experiments have played an important role in theorizing about psychodynamics. White (1948) among others has leaned extensively on the Diven experiment in his theory of anxiety and the learning of neuroses.

There are a number of methodological flaws in both the Diven and Haggard experiments that make the conclusion of unconscious conditioning of anxiety at best very tenuous. Several of these defects have been pointed out by Lacey and Smith (1954) who repeated the Diven experiment with an improved design.

They used a somewhat different word list and two experimental groups. *Ss* in the Cow-shock group were given shocks following their associations to the word Cow while *Ss* in the Paper-shock group were shocked following their associations to the word Paper. Heart rate rather than GSR was used as the conditioned response, and the anticipatory changes in heart rate that occurred during the *S*'s associations were analyzed instead of the responses following association, as was the case in the Diven and Haggard studies. These changes in method, however, appear to have had little effect since Lacey and Smith conclude that not only did both groups of *Ss* show conditioning of anxiety reactions without awareness but also generalized the conditioned response to semantically similar words.

In spite of their changes in design, the Lacey and Smith study still lacks some crucial controls that would make the interpretation of unconscious autonomic conditioning unequivocal. Eriksen (1958) has pointed out that their method of determining awareness in *Ss* and their method of computing conditioning are such as to permit a spurious precision in the autonomic conditioning relative to the verbal awareness.

They determined awareness by intensively questioning *Ss* as to whether they had learned when to expect shock. *Ss* classified as unaware were those who were unable to verbalize the specific contingent relationship between shock and the word Cow (Paper). This method of questioning, however, may not elicit all the information available to the *S*'s verbal report. It is quite possible that the *Ss*, if asked, could have given several or more words which they were quite confident were not shocked. It seems reasonable to assume that the *S*'s change of heart rate to a stimulus word which he is confident will not be followed by shock will be less than to those words of which he is still unsure (potentially dangerous). It would also seem reasonable to expect the actual shock word in general to be among those words that the *S* was uncertain of rather than among the words he was rather confident were not followed by shock, while a semantically dissimilar word would more likely be in the latter category.

The evidence that previous studies have found on GSR generalization to words in the same semantic classification might actually represent not generalization so much as a beginning on the part of the *S* to learn to dis-

[1] This research was supported by Mental Health Grant M-1206, National Institutes of Health, Public Health Service.

criminate the general area where shock occurs. Thus, *Ss* in the Cow-shock group would tend to classify rural words in the uncertain category and nonrural words including paper in the confident not shock category. The reverse would occur for the paper shock *Ss*. In view of this possibility it becomes apparent that the control measure used to determine whether conditioning is occurring becomes quite crucial.

In the study by Lacey and Smith, conditioning was computed by taking each S's heart rate score to cow and subtracting from it his heart rate score to the word paper. It should be readily apparent that this method of computing conditioning is apt to lead to a spurious appearance of precise discrimination by the heart rate response which in fact may not exist. To show that conditioning of heart rate is more specific than the stage of the *S*'s verbal awareness, an analysis is required that demonstrates whether or not the *S* shows greater autonomic conditioning to the actual shock word than he does to the rest of the words concerning which he verbalizes uncertainty as to whether they were followed by shock.

The experiment reported below replicates the Lacey and Smith study in its essential details with the exception that *S*'s verbal awareness was explored more extensively. The *S*'s verbalizable expectancies concerning shock were obtained for each stimulus word in order to control for the possible artifact described above. In addition, some of the characteristics of the experimental situation that were related to the *Ss*' expectancies were explored.

METHOD

The procedure was administered to each *S* individually. *Ss* were comfortably seated with the experimental apparatus screened from their view and were instructed that this was an experiment on the effects of tension on mental associations. They were further told that a list of stimulus words would be presented to them one at a time by means of a tape recorder, and upon the occurrence of a stimulus word, they were to chain associate to this word until the stop signal occurred. The process of chain association was explained to them, and then they were given practice in chain association to three nonexperimental words to insure that they had adequately comprehended the directions.

Next, the shock electrodes were connected to the right forearm of the *S*, 2 in. above the wrist with an interelectrode distance of 1 in. The electrodes were fixed firmly into position by masking tape. The *S* was told that the shock was an integral part of the experiment designed to produce tension on his part and would occur several times throughout the word list. At this point, the *S* was given several sample shocks in order to adjust the voltage level of the shock to the point where the *S* reported it as very painful. The GSR was recorded from the dorsal and ventral surface of the middle finger of the left hand through silver electrodes. The finger was cleaned with alcohol and treated with electrode jelly. GSR records were recorded continuously throughout the experimental session.

Upon completion of the instructions and after a short period of adaptation, the experimental session began. The tape recorder presented the stimulus words in the same order as that employed by Lacey and Smith. The words had been recorded by a person trained in speech to minimize distortion or confusability in pronounciations. Following the occurrence of the stimulus word, 15 sec. were allowed for the *S*'s chain associations before the word stop occurred. Fifteen seconds then elapsed between the stop signal and the occurrence of the next stimulus word. The 40-item stimulus list consisted of the words, Cow, Paper, Easy, and Harvest six times each, eight other rural words and eight nonrural words (common adjectives and nouns) all distributed in temporally balanced order in the list.

During the conditioning session (one presentation of the 40-item list), one group of *Ss* (Cow-shock) was administered a painful shock coinciding with the stop signal following each presentation of the word Cow. The other group of *Ss* (Paper-shock group) were shocked similarly following the word Paper.

Upon completion of the conditioning phase, the following questions were asked and the *S*'s answers recorded verbatim:

1. Did you know when you were going to get shocked?
2. Please explain freely what you think about the shocks in this experiment.
3. Please try to explain to me what you think or guess is the scheme behind the administration of the shock. Try to guess in any manner you like—you can guess as much as you like, all I'm interested in is what you think about or make out concerning shock.
4. Let me read the list of stimulus words and let us see whether you can recognize any particular word or words which might have been followed by shock. I want to see whether you can distinguish between words which might have been followed by shock and those which might not have been followed by shock. You can say No, Yes, Uncertain or whatever you think concerning the probability of the particular word having been followed by shock. If you're not sure, guess.

An extinction phase followed the inquiry. It was identical to the conditioning session with the exception that no shocks were administered although the electrodes were left undisturbed and the *Ss* were not informed that there would be no further shocks.

Apparatus. Electric shock was administered by means of a Harvard inductorium using a faradic stimulation. The primary coil was supplied from a 3-v. dc source. The voltage of the secondary was

adjusted for each *S* to a level that was reported very painful.

The GSRs were continuously recorded with a modified Deceptograph supplied by C. H. Stoelting and Company. The modification was such as to permit the continuous reading of the *S*'s basal resistance level in ohms.

Subjects. The 22 *S*s consisted of 4 men and 18 women who were volunteers among graduate students in the College of Education at the University of Illinois. They were randomly assigned to the Cow-shock and Paper-shock groups. None was familiar with the original Diven experiment nor the experiments of Haggard and Lacey and Smith. The Cow-shock group contained 12 *S*s and the Paper-shock group 10.

RESULTS

On the basis of the *S*s' verbalizations during the inquiry, they were classified into aware and unaware *S*s. Three *S*s in the Cow-shock group and one in the Paper-shock group were able to correctly verbalize the contingent relation between shock and the correct stimulus word. The remaining 18 *S*s were classified as unaware. Their verbalizations were vague concerning the stimulus word followed by shock, although some verbalized incorrectly but with confidence that particular words had been followed by shock one or more times during the conditioning session. The four aware *S*s have been excluded from the following analyses.

The first step in the analysis was to determine the extent to which the present data duplicated the findings of Lacey and Smith. Complete GSR records were available on seven unaware *S*s in each of the two shock groups.[2] The maximum GSR deflection in micromohs of conductance change that occurred during the 15-sec. association interval was determined for each of the 40 stimulus words for each *S*. The list of stimulus words was then divided into four blocks of ten consecutive words each and the mean and SD of the GSR deflections for each *S* for each of these four blocks was computed. GSRs to the different words were then converted into *z* scores based upon the mean and SD of the 10-word block in which they occurred. These transformed GSR *z* scores were used in all the subsequent analysis.

To render our data comparable with those of Lacey and Smith, their method of determining conditioning was first computed. In this method, the transformed GSR *z* score to paper (in their study, heart-rate) is subtracted from the response to Cow for each *S*. Since words in corresponding positions are taken for comparison, any temporal trend is controlled. The technique of moving averages by taking the mean of consecutive blocks of three trials is used to smooth out irregularities in the conditioning curves for the two shock groups.

The results are graphed in Fig. 1. In terms of this measure of "conditioning" the Paper-shock group gives evidence similar to that obtained by Lacey and Smith with the maximum GSR occurring on the sixth block of trials, following which extinction takes place. For the Cow-shock group, however, there appears to be no consistent evidence of conditioning until approximately the seventh trial which is in the extinction period and after the pause for questioning. By the tenth trial, however, there is some evidence of extinction in this group.

Lacey and Smith evaluated the significance of these trends by the nonparametric one sample runs test applied to the response differences block by block. Applying this test to our data, we find that only in the seventh block was a difference between the two groups sig-

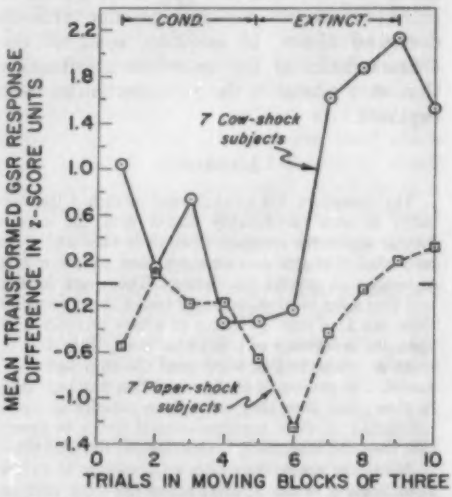

FIG. 1. Conditioning curves computed by the method of Lacey and Smith (1954).

[2] Due to recording difficulties, the GSR records for two *S*s in each group were not complete. These four *S*s were excluded from the analyses of the GSR data.

TABLE 1
MEAN TRANSFORMED GSR READINGS FOR THE SHOCK GROUPS FOR DIFFERENT GROUPINGS OF WORDS DURING THE CONDITIONING SESSION

Shock-group	Shock Word	Repeated Similar Words	Repeated Nonsimilar Words	Nonrepeated Similar Words	Nonrepeated Nonsimilar Words
Cow-shock group	0.38	−0.044	−0.42	1.81	−0.93
Paper-shock group	1.82	1.37	−1.33	2.29	−0.58

nificant at the .05 level (one-tailed). However, summing for each *S* through the six conditioning trials and testing the mean difference between the Cow-shock and Paper-shock group resulted in a t value of 1.30, which with 12 degrees of freedom is not statistically significant. Summing through eight trials (taking the six conditioning trials and the first two extinction trials together) and then testing the mean difference for the two shock groups yields a t of 2.83, which is significant beyond the .01 level (one-tailed). A similar t test applied to the sum of the six extinction trials resulted in a value of 2.31, which again with 12 degrees of freedom is significant beyond the .03 level (one-tailed).

Evidence of conditioning within each of the two groups during the conditioning phase was evaluated by a three-way analysis of variance (trials, Cow and Paper, *S*s) of the z scores to Cow and Paper. A separate analysis was performed for each group. In these analyses conditioning to the shock word would be reflected in a significant interaction between trials and the Cow and Paper words.

Only in the Paper-shock group was there evidence of conditioning with the interaction term significant at approximately the .05 level ($F = 2.50$ with 5 and 30 *df*). The main effects did not approach significance in either group and the absence of significant interaction in the Cow-shock group provides no evidence of conditioning for these *S*s.

Lacey and Smith also found that the conditioned responses tended to generalize in the respective groups to words of a similar semantic classification. Thus, the Paper-shock group showed a greater response to nonrural words than to rural words, whereas the Cow-shock group gave greater responses to the rural words than to the nonrural words. The present data were examined for evidence of this type of generalization. In Table 1, the mean transformed GSR readings for the two experimental groups during the conditioning session are shown as a function of different groupings for the words. As is seen, the mean GSR for the nonrepeated similar words in both groups is higher than that for the actual shock words, a finding which is in conformity with that of Lacey and Smith who also noted that the generalized response magnitude tended to be higher than the response to the actual conditioned stimulus.

The significance of the differences observed in Table 1 was evaluated by a modified three-way classification analysis of variance (groups, word classification, and individuals). The word classification was broken down into shock words, repeated similar words, repeated nonsimilar words, nonrepeated similar words, and non-repeated nonsimilar words. The results of this analysis failed to show any significant effects, with the F ratio for between word classifications only 1.54 (4 and 48 *df*, $p > .05$).

Since, as will be seen below, repetition of words in the list has a confounding effect upon any measure of conditioning or awareness, a more meaningful test for generalization is in terms of the nonrepeated rural and nonrepeated non-rural words in the list. Again a modified three-way classification analysis of variance similar to the above was performed but in this case only two word classifications were used, nonrepeated rural and nonrepeated nonrural words. None of the main effects in this analysis was significant, but the interaction between the two shock groups and rural-nonrural word classification was significant beyond the .05 level ($F = 5.87$ with 1 and 12 *df*). This significant interaction reflects the trend that can be seen in the above table where it is found that the Paper-shock group shows greater GSR to nonrural words while the reverse effect is true for the Cow-shock group. Similar analysis of variance performed upon the extinction trials failed to yield any significant effects.

The data analyzed so far are in general agreement with the results presented by Lacey and Smith. For the Paper-shock group, at least, they could be interpreted in terms of rather

TABLE 2
LEVEL OF CERTAINTY OF SHOCK SCALE AND CORRESPONDING SCALE SCORES

Scale Score	Scale Category	Verbal Responses included in Each Scale Category
6	Yes	Yes
5	Once or twice	Maybe once or twice, Once, Once or twice
4	Maybe	Maybe, Perhaps, Possibly, Could be, Probably, I think so, I believe so
3	Uncertain	Uncertain, Not sure
2	I don't know	I don't know, I don't remember, Never heard, Didn't hear, Can't see any connection
1	I don't think so	I don't think so
0	No	No

specific autonomic conditioning in the absence of verbal awareness. However, as was previously pointed out, this evidence for specificity of conditioning might very well be an artifact of the method of computing or analyzing conditioning. The analysis that has been conducted so far has not determined verbal awareness in the *S*s in a manner that is commensurate with the analysis of autonomic discrimination or conditioning. While all of the *S*s contributing to the data were unaware in the sense that they could not verbally report that shock followed Cow (Paper), they nonetheless had various expectancies concerning the relations between certain words and shock. The existence of these expectancies was revealed by the more detailed assessment of awareness that we employed in this study.

As will be recalled, we presented each of the 20 different stimulus words to the *S* and asked him to state whether or not these particular words had been followed by shock. It was possible to sort their replies on a seven-point scale defining the expectancy level of the *S* as to whether a word had been followed by shock. Table 2 shows how the Ss' replies were classified in terms of the scale scores.

For the *S*s in the Paper-shock group, the average number of words per *S* receiving a rating of three or higher was 5.4 (range 3–8). For the *S*s in the Cow-shock group, the mean was 6.8 (range 5–12). It was further found that five of the seven *S*s in the Paper-shock group had given a value of three or higher to the word paper, and four of the seven *S*s in the Cow-shock group had given a value of three or higher to the word Cow. The words receiving scale values of three or higher may be considered to be perceived by the *S*s as potentially dangerous relative to the words having scale values that are lower than three.

Our next step was to determine whether the GSR response was greater to the actual shock word than it was to the other words that *S* perceives as potentially dangerous. For each *S*, the GSRs in *z* score units during the conditioning trials were averaged for all words to which he had given a rating of three or higher, excluding the actual shock word. The difference between this mean GSR to the words perceived as dangerous and the GSR to the actual shock word was computed for each *S*, and the mean of the differences by experimental group was evaluated with a t test to see whether it differed significantly from zero. For the Paper-shock group, the mean of the *z* scores for the shock word was .18 as was compared with a mean of .26 for the words with scale values of three or more. The mean difference of .08 although not significantly different from zero ($t = .09$) is actually in the opposite direction from what would be expected if the *S* were unconsciously discriminating the true shock word from the other words perceived as dangerous. For the Cow-shock group, the comparable means were .06 and −.16. Again the mean difference (.22) although now in the direction for unconscious conditioning, is not significantly different from zero ($t = 1.54$, $p > .05$). It would appear therefore, that in the present experiment the autonomic conditioning manifested by the *S*s is no more specific or precise than their verbally stated awareness.

It is possible now to show not only that the evidence given by the Paper-shock group for autonomic conditioning without awareness is artifactual, but also why the Cow-shock group did not show such evidence. Examination of the *S*'s expectancy of shock for the different words shows that only two of the seven *S*s in the Paper-shock group report having been shocked to the word *Cow* while five of the seven *S*s in the Cow-shock group report having been shocked to *Paper*. Since the GSR tends to be greater to words which *S*s report as having been followed by shock, assessing conditioning for the *S*s in the Paper-shock group by sub-

TABLE 3
KENDALL'S COEFFICIENT OF CONCORDANCE, W, OBTAINED FROM DISTRIBUTION OF RANKED LEVEL OF CERTAINTY OF SHOCK FOR DIFFERENT GROUPINGS OF SUBJECTS AND STIMULUS WORDS

Grouping	No. of Ss.	No. of Words	Kendall's Coeff. (W)	Significance (p)
Cow-shock group: Rural words only	9	10	0.127	.30
Cow-shock group: Nonrural words only	9	10	0.160	.15
Cow-shock group: All words together	9	20	0.129	.20
Paper-shock group: Rural words only	9	10	0.255	.01
Paper-shock group: Nonrural words only	9	10	0.468	.01
Paper-shock group: All words together	9	20	0.360	.01

tracting the autonomic response to paper from the autonomic response to Cow would give the general results shown in Fig. 1.[3] Also, the fact that a high number of Cow-shock *Ss* report having been shocked to Paper would lead us to expect negative results from this type of subtractive computation.

It is informative to analyze the expectancies of the *Ss* further in order to identify features and characteristics of the experimental arrangement that lead *Ss* to identify certain of the stimulus words as being associated with shock. First, we examined the consistency among *Ss* within experimental groups in the assignment of their expectancies to different stimulus words. For this purpose, Kendall's coefficient of concordance (W) was employed. Table 3 shows the Ws for rural words, nonrural words, and all words together between *Ss* in each of the two experimental groups.[4]

As can be seen from the table, the *Ss* in the Cow-shock group did not show a significant amount of consistency in their expectancies for the different words together or when broken down into separate categories. The Paper-shock group on the other hand showed significant consistency in all three analyses. Thus, only in the Paper-shock group is there evidence that the administration of shock had a systematic effect upon the hypotheses the *Ss* construct concerning shock.

Despite lack of consistency among the Cow-shock *Ss* there is actually considerable consistency in the assignment of certainty to words between the two shock groups. If the average expectancy for each of the 20 different stimulus words is computed separately for the two experimental groups, the rank order correlation of these expectancies between groups for the 20 words is .62 ($p < .05$). This finding would suggest that there are some elements in the experimental arrangement that biased the *Ss*' ideas concerning the occurrence of shock.

To identify some of these factors, the stimulus words were subdivided into five categories: the shock word; repeated similar words; repeated dissimilar words; nonrepeated similar words; and nonrepeated dissimilar words. A modified three-way classification analysis of variance (groups, word classification and individuals) were performed on the expectancy scores for the different words. This analysis revealed a significant variation between word classification ($F = 6.81$ with 4 and 64 df, $p < .01$), but there were no significant differences between the two groups nor was the interaction between groups by word classification significant. Inspection of the cell means from this analysis suggests that the significant effect due to word classification is attributable to the tendency for words that have been repeated six times in the list to have higher shock expectancy than words which have occurred only once, and, further, that words belonging to different semantic categories are also being discriminated with regard to expectancy. Apparently, the *Ss* have begun to differentiate at the verbal level the experimental situation in terms of two relevant characteristics: the repetition of certain stimulus words and semantic classification.

[3] This method of assessing conditioning for the two *Ss* in the Paper-shock group who reported having been shocked to Cow shows little or no evidence of conditioning. This result is to be expected in view of the fact that they had also included Cow as a potentially dangerous word in their verbal reports.

[4] For these analyses, N in each group is nine since the expectancy data were obtained for two *Ss* in each group on which complete GSR data was not available.

DISCUSSION

Lacey and Smith (1954) in their penetrating analyses of the Diven (1937) and the Haggard (1943) experiments have quite accurately pointed out that neither of the experiments contained the necessary controls to support a conclusion of "unconscious conditioning." Diven compared the GSR of the experimental *Ss* to the shocked word with that of a control group who had received no shocks at all. As Lacey and Smith (1954) point out, this comparison is invalid in the absence of a demonstration that the experimental group's reactivity to all words was not greater than the control group's. This criticism also applied to Haggard's study. And as was seen above, while Lacey and Smith use a more sophisticated and appropriate statistical test of conditioning, they still have not excluded the likelihood that their evidence of "unconscious conditioning of anxiety" is artifactual.

The methodological flaw in the Lacey and Smith study is the same one that has plagued a number of attempts to demonstrate learning without awareness. Adams (1957) has termed it the problem of correlated hypotheses and has shown how failure to consider this problem has invalidated many of the experiments in this area.

The problem arises because these experiments focus on and record only a limited aspect of the *Ss*' behavior. When the *S* is examined for awareness, he is classified as unaware if he is unable to repeat the verbalization the *E* has formulated. But his over-all behavior in the experiment may be in perfect correspondence with his verbalizations. What the experimenter overlooks is that the change or modification in the observed response could actually be correlated with a number of different hypotheses. Convincing research in learning without awareness must of necessity thoroughly explore the *Ss* verbalizable hypotheses and expectations and only when it is shown that these hypotheses are unable to account for the observed change in the response is a conclusion of learning without awareness warranted.

At present we are left with no convincing evidence that specific autonomic conditioning occurs in the absence of verbal awareness. Razran (1949) has reported salivary conditioning in human *Ss* without their awareness, but as Lacey and Smith point out, Razran's description of his method of determining awareness is so sketchy as to leave serious doubts as to whether he successfully avoided the pitfalls that have been discussed.

We would suggest that the future investigations in this area would do well to abandon the method originated by Diven (1937). Our results suggest that there are so many uncontrolled and confounded sources of variance in this design as to render virtually impossible an answer to the basic question of conditioning without awareness. Both the *S*'s expectancies of shock as well as his GSRs were a function of the frequency of repetition of certain words in the list. The introduction of new words toward the end of the list also seemed to influence the GSR, perhaps as a result of their novelty or unexpectedness. In addition, the 15-sec. CS–US interval is quite long for conditioning to occur, since the optimum interval is generally considered to be much shorter.

Generalization of the autonomic response would more appropriately be measured by the introduction of new words after the conditioning trials have taken place. The present design results in a certain amount of masking of what generalization effects there are due to the fact that the *S* is presented the generalization stimuli interspersed with the conditioning trials and thus is learning a condiioned discrimination.

There is one other aspect of our data that deserves brief comment. This is the increase in GSR to the conditioned stimulus in the Cow-shock group during the extinction trials, (Fig. 1), an effect which is particularly interesting since the *Ss* in this group gave no evidence of conditioning during the six conditioning trials. It is impossible to evaluate the meaning of this effect since, unfortunately, we did not requestion the *Ss* following the extinction phase of the experiment. What changes may have occurred in their expectancies concerning shock as a result of the inquiry period can only be surmised.

SUMMARY

This study was essentially a replication of the Lacey and Smith (1954) experiment on the unconscious conditioning of autonomic responses. When their methods of data analysis were used, their results in general were con-

firmed. However, due to a more extensive and careful investigation of the *Ss'* awareness in the present study, it was possible to show that the conditioning of the autonomic response (GSR) was no more specific than the *Ss'* verbalizations. It was also shown that the method employed by Lacey and Smith to compute conditioning gives rise to a spurious conclusion of precise autonomic conditioning without verbal awareness. The limitations of their basic design for investigation of the problem of unconscious conditioning were pointed out.

REFERENCES

ADAMS, J. K. Laboratory studies of behavior without awareness. *Psychol. Bull.*, 1957, **54**, 383–405.

DIVEN, K. Certain determinants in the conditioning of anxiety reactions. *J. Psychol.*, 1937, **3**, 291–308.

ERIKSEN, C. W. Unconscious processes. In M. R. Jones (Ed.), *Nebraska symposium on motivation*. Lincoln: Univer. Nebraska Press, 1958.

HAGGARD, E. A. Experimental studies in affective processes: I. Some effects of cognitive structure and active participation on certain autonomous reactions during and following induced stress. *J. exp. Psychol.*, 1943, **33**, 257–284.

LACEY, J. L., & SMITH, R. L. Conditioning and generalization of unconscious anxiety. *Science*, 1954, **120**, 1045–1052.

RAZRAN, G. Stimulus generalization of conditioned responses. *Psychol. Bull.*, 1949, **46**, 337–365.

WHITE, R. W. *The abnormal personality*. New York: Ronald, 1948.

(Received July 20, 1959)

Journal of Abnormal and Social Psychology
Vol. 60, No. 3, 1960

THE EFFECTS OF SUBLIMINAL AND SUPRALIMINAL SUGGESTION ON VERBAL PRODUCTIVITY[1]

MARVIN ZUCKERMAN

Institute of Psychiatric Research, Indiana University Medical Center, Indianapolis

INVESTIGATIONS of "subliminal" influences on behavior are not new in psychology; studies in this area date back to the 19th century. However, judging from recent reviews of the literature by Adams (1957), McConnell, Cutter, and McNeil (1958), Eriksen (1958), and Goldiamond (1958), a resurgence of interest in the area is underway.

Most of the studies reviewed by Adams (1957) require a subject to make a stimulus discrimination. In one group of these experiments, the subject knows the nature of the cue but his discrimination is better than chance at stimulus intensities at which he thinks he is guessing. In another group of experiments the subject does not know that a cue to discrimination is being given. Only 9 of the 57 studies in Adam's review do *not* involve a stimulus discrimination task. Seven of these are essentially operant conditioning studies in which the subject does not realize he is being conditioned because he is unaware of either the reinforcer or the class of responses being reinforced, or of both. The use of subliminal suggestion in movie and television advertising seems to belong with this latter group, for the subject's operant behavior is supposedly brought under control by suggestive cues of which he is not aware. The present study represents an attempt to construct an analogue of this type of situation. The question asked in the first experiment was: "Can one influence frequency of a graphomotor response by suggestions administered visually at subliminal intensities?"

EXPERIMENT I

The experimenter tried to control the *S*'s verbal productivity (the number of words written in describing pictures) with two subliminal suggestions: WRITE MORE and DON'T WRITE. A control group of *S*s described the pictures without subliminal suggestion. Each of these suggestions was typed in capital letters on a paper-roller fitted into a tachistoscope. Three TAT cards—4, 6BM, and 7GF—were used as the pictures to be described in writing. These pictures, which each portray two persons with one turned away from the other, were rotated over the conditions in the experiment so that each picture appeared an equal number of times in each condition.

Subjects

The 36 *S*s consisted of 33 student nurses and three secretaries. Their average age was 20.0 (SD = 1.3). The student nurse *S*s were paid volunteers.

Apparatus

The apparatus used was the Harvard Tachistoscope, a modification of the Dodge Mirror Tachistoscope. This model, manufactured by Ralph Gerbrands Co., has dual timers permitting control of two stimulus fields. Its fluorescent lamps may be turned on and off instantly. Dual mechanical timers provide exposure intervals of .01 to 1 sec. for two sets of lamps, each set controlling a field of exposure. The duration of the overlap or interval in the presentation of the fields is controlled by a third timing mechanism. Single pieces of Atlantic bond paper (20 wt.) were used as filters on the two lamps controlling the second exposure field.

Procedure

The *S*s were told that they would be shown three pictures in the tachistoscope and that each picture would be shown 10 times, each time for a slightly longer interval than the preceding time. After each presentation, they were to write on a pad everything they saw in the picture, including everything that was happening in the picture. They were told that it was permissible to repeat what they had written on previous presentations of the same picture but they were told not to use ditto marks or words like "same as before."

Before starting the experiment the word "SAMPLE" was presented for .02 sec. against a blank, lit background field. The *S* was asked if she had seen either a word or letters presented. If (as was the case with two *S*s) she did report a word or letters, one additional bond paper filter was inserted in front of each of the two lamps controlling the second exposure field, which contained the word; to insure that the experimental words would be subliminal. All but two *S*s reported

[1] James A. Norton, Department of Mathematics and Statistics, Purdue University, was the statistical consultant for this study. He devised the modifications of standard designs and tests of significance used in the analyses of the data.

seeing merely a "flash," a "blur," or, in a few cases, "lines."

Condition I was the same for both experimental and control *Ss*. The first exposure field, containing a TAT picture, was exposed ten times for successive durations beginning at .05 sec. on the first trial and lengthening .05 sec. on each successive trial, reaching .5 sec. on the 10th trial (ostensibly, the experiment was to test the effect of different durations of presentation on perceptual acuity). The second exposure field was blacked out except for an exposure slot in which the stimuli were to appear during the experimental conditions. During Condition I this slot was blank, resulting in a flash of .02 sec. duration overlapping with the last .02 sec. of the presentation of the TAT picture and superimposed on the picture. After each trial the *S* was allowed to write her description of the picture. She was instructed to say "ready" whenever she finished writing. No time limits were placed on the *S*. However, because of the close scheduling of *Ss* it was necessary to discard *Ss* who took longer than half an hour to complete the ten trials of Condition I. Three *Ss* were discarded because of this limitation. As soon as the *S* signalled that she was ready for the next presentation, the examiner gave a ready signal and pressed the timer button.

In Condition II, a second TAT card was presented. *Ss* followed the same procedure as in Condition I. The *Ss* in the control group again had a blank in the second exposure field. However, the *Ss* in the experimental group were given the stimulus words "WRITE MORE" in the second exposure field on each of the ten trials. This stimulus overlapped for .02 sec. with the last .02 sec. of the picture being presented in the first field. It appeared approximately in the middle of the picture.

In Condition III a third TAT picture was presented with the *Ss* following the same procedure. The *Ss* in the control group again had a blank in the second exposure field. The *Ss* in the experimental group were given the stimulus words "DON'T WRITE" in the second exposure field, again overlapping for .02 sec. with the end of the exposure period of the picture.

After the last trial on the last condition, a blank card was inserted in the first exposure field which now served as a lit background field. The stimulus words from the last condition were presented at the .02 sec. setting and the *S* was asked to describe what she saw. The stimulus presentation was lengthened .01 sec. on each successive trial until the *S* could correctly read it. This threshold was obtained in order to determine how close to threshold the *S* was functioning during the last condition of the experiment. The mean threshold for the 18 *Ss* in the experimental group was .08 sec. (*SD* = .04). However, it was discovered in the second experiment to be reported that it is much more difficult for *Ss* to recognize words presented at the end of a primary stimulus to which they are attending, than it is to recognize the stimulus against a blank background. It is likely that the actual threshold for the words against the background of a complex picture was much higher than the threshold obtained against a blank background. However, on the assumption that this measure might have some correlation with the real threshold, it was examined in relation to the magnitude of the suggestion effect.

TABLE 1
EXPERIMENTAL DESIGN

Group and Card Order	*N*	Conditions		
		I	II	III
Control				
ABC	6	.02-sec. blank	.02-sec. blank	.02-sec. blank
BCA	6			
CAB	6			
Total	18			
Experimental				
ABC	6	.02-sec. blank	.02-sec. words ("Write More")	.02-sec. words ("Don't Write")
BCA	6			
CAB	6			
Total	18			
Total *Ss*	36			

After the threshold was obtained, *Ss* were carefully questioned as to whether they saw any words being presented with the pictures. Only one *S* in these two groups reported seeing the stimulus words and she was discarded as a *S* in the experiment.

Ss were urged not to tell any details of the experiment to other nurses. Actually almost all of them were still unaware of the nature of the experiment. Most of them seemed surprised when they were asked if they noticed words presented with the pictures. New *Ss* were questioned as to what they had heard about the experiment and as far as could be determined, none of them had heard that words were presented along with the pictures.

Design

The design is presented schematically in Table 1. The design was one of replicated 3 × 3 latin squares. The complete analysis of this design is an extension of the analysis presented by Edwards (1950, p. 326).

Results

The "Residual within Individuals" sum of squares in the design is made up of six parts: a Subjects × Conditions interaction (with 10 *df*) from each of six Group × Order of Presentation combinations. These six contributions were computed separately and tested for homogeneity with the Bartlett test, which yielded a chi square of 15.09 (significant at the .05 level). Since the variance was heterogeneous a logarithmic transformation was employed. Bartlett's test was applied

TABLE 2
ANALYSIS OF VARIANCE

Source of Variance	*df*	Mean Square	*F*
Between Groups	1	.298	2.54
Between Orders	2	.203	1.73
Groups × Orders	2	.023	.20
Between *Ss* within Groups × Orders	30	.117	
Between Conditions (avg. both groups)	2	.101	12.64**
Conditions × Groups	2	.027	3.44*
Between Cards	2	.001	.16
Cards × Groups	2	.013	1.66
Latin Sq. Residual	2	.006	.75
Latin Sq. Residual × Groups	2	.030	3.81*
Residual within *Ss*	60	.008	
Total	107		

* $p = .05$.
** $p = .001$.

TABLE 3
GROUPS MEANS (LOGS) ON CONDITIONS

	I	II	III
Control	2.11	2.15	2.24
Experimental	2.22	2.31	2.29

to the transformed data and the resultant chi square of 10.43 was not significant at the .05 level.

Table 2 presents the results of the complete analysis of variance. For the central hypothesis only the Conditions × Groups interaction is relevant. The effects of Cards and Order of Presentation of the Cards are only of secondary interest.

The Groups × Conditions interaction mean square was .02736, giving an F of 3.44 with 2 and 60 *df*, significant at the .05 level. Thus, the subliminal stimulus conditions did result in a difference betweeen groups. The nature of this difference can be seen in Table 3 which gives the means of the two groups (in logarithm scores) for the three conditions. The control group showed a steady rise in productivity going from Conditions I to III. The experimental group showed a slightly greater rise in productivity from Condition I (blank) to Condition II (WRITE MORE) and showed a drop instead of a rise from Condition II (WRITE MORE) to Condition III (DON'T WRITE).

In order to examine more closely the effect of the subliminal stimulus, individual comparisons were carried out on the difference in mean change from one condition to another in the two groups.[2]

Three t tests were performed between the three Conditions. Only the difference between groups from Condition II to Condition III was significant ($t = 2.61$, $p = .025$ for a one-tailed test). Apparently the subliminal suggestion "DON'T WRITE" played the greater role in the interaction between groups and conditions. On Condition III, 13 of the 18 *Ss* in the experimental group decreased their productivity (in accord with the subliminal suggestion) while only 4 of the 18 *Ss* in the control group decreased their productivity.

Condition III consisted of 10 trials. On each successive trial the picture stimulus was exposed for a longer interval. One might expect the subliminal suggestion to have some kind of increasing effect as trials progress. Figure 1 contains plots of the mean and median (logarithms) numbers of words written on each trial for the two groups. There was considerable variability within groups on single trials causing the means to be somewhat distorted by extreme results. In terms of mean scores, it can be seen that the experimental group started on Trial 1 at a considerably higher level of productivity. This difference is probably a carry-over from Condition II. The large initial difference between the groups on Trial 1 disappears by Trial 3 as a result of the rapid increase in the Control group and the decrease in the Experimental group. Beyond this trial the groups do not markedly diverge. The medians present a somewhat different picture. The experimental group remains more productive up to the seventh trial, and from the seventh through the tenth trial drops below the control group. With the exception of an "end-spurt" in both groups, there is a tendency for the experimental group to *decrease* while the control group *increases* over the successive

[2] For instance, for the change from Condition II to III we have:

$$t = \frac{(\bar{Y}_{eII} - \bar{Y}_{eIII}) - (\bar{Y}_{cII} - \bar{Y}_{cIII})}{\sqrt{MS_{w/s}}} \text{ with } 60\ df$$

where $\bar{Y}_{eII}$ is the mean for the experimental group on Condition II, etc. These t's were judged for significance using Tukey's Studentized Range test for multiple comparisons (Scheffe, 1953), the .05 level in this case being 2.40.

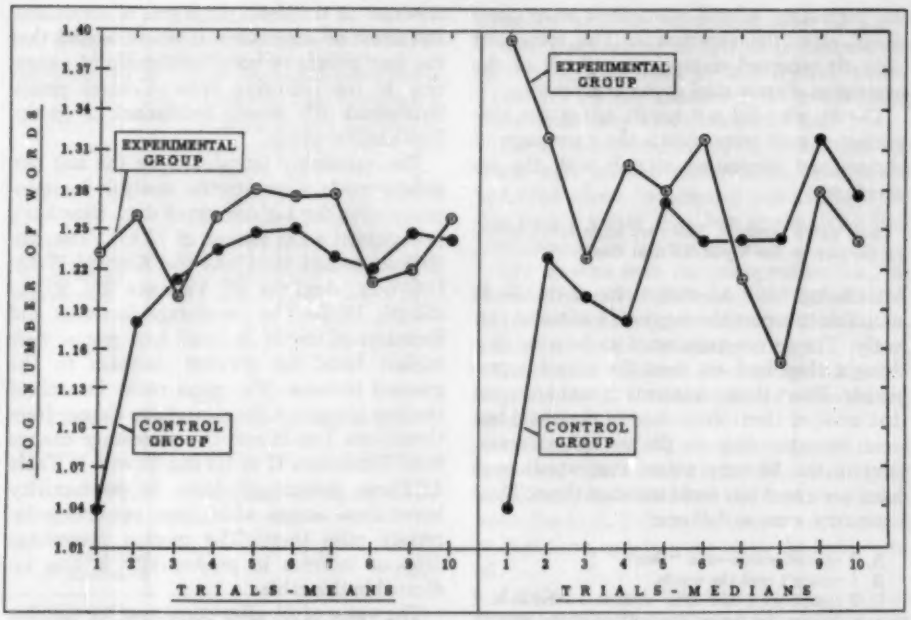

FIG. 1. Means and Medians (logarithms) in the Experimental Group on the Ten Trials of Condition III ("DON'T WRITE").

trials. Thus, there is some indication that the subliminal suggestion "DON'T WRITE" has a cumulative effect, when presented over a series of trials.

In order to examine the relationship between the *S*'s threshold and her response to the subliminal suggestion, a rank order correlation was performed between the percentage drop in productivity on Condition III in the Experimental group and the *S*'s threshold for the words "DON'T WRITE" obtained against a blank background at the conclusion of the experiment. The 18 experimental group *S*s were ranked from the greatest percentage increase to the greatest percentage decrease in productivity on Condition III, and thresholds were ranked from the highest to the lowest. The rho obtained between these two rankings was −.32. Although this value falls short of significance, the direction of the correlation suggests that high thresholds tended to be associated with low, or weak, response to the subliminal suggestion.

The results up to this point indicated that one of the subliminal suggestions had a significant influence on the productivity of the *S*s. The direction of the correlation between the threshold and the magnitude of this influence suggested that, perhaps, the influence of the suggestion increases as the *S* becomes more aware of it. A third group of *S*s was therefore tested, this time increasing the duration of the suggestion stimuli so that they would become supraliminal.

EXPERIMENT II

Method

Eighteen more student nurse *S*s composed the third group of the experiment. Their mean age was 19.94 (*SD* = 1.30). The procedure was the same as that of the previous experimental group except that this time the duration of the suggestion stimulus was increased to .5 sec., its onset beginning .02 sec. before the shut-off of the picture stimulus. It was assumed that this half-second duration would make the suggestion stimulus supraliminal for all *S*s. The threshold for the Condition III suggestion stimulus presented against a blank background had not exceeded .2 sec. for any of the *S*s in Experiment I.

Results

To the surprise of the experimenter, 10 of the 18 *S*s still did not report seeing either of

the suggestion stimuli when they were questioned after the experiment. The remaining eight *S*s reported seeing one or both of the suggestion stimuli, during the experiment.

The *S*s who did not report seeing the suggestion stimuli were shown the succession of picture and suggestion stimuli with the instructions:

> Now watch carefully. There are some words following the picture. See if you can read them.

After being "set" all except one of the 10 *S*s were able to report the suggestion stimulus correctly. They were questioned as to why they thought they had not seen the stimulus previously. From their comments it was apparent that most of them were close to threshold but were not attending to the second stimulus. Four of the 10 were aware that words were being presented but could not read them. Their comments were as follows:

> A. I noticed words—saw "Don't."
> B. I couldn't read the words.
> C. I remember I saw little squares. . . blur in it. I thought it was the tag on the bottom of the picture. If I'd been paying attention I could have seen it.
> D. It looked like it might be the title of the picture. . . . Printing in the center of the card I was looking at.

Three *S*s were aware of only the flash. Their comments were:

> E. I noticed a flash. I guess it was because I was concentrating so hard on the picture. . . . I was annoyed at the light. . . . Just trying to get it out of the way more or less.
> F. I saw a flash but it was too quick.
> G. Flash . . . but I was just expecting pictures.

The remaining three *S*s did not notice anything but the pictures.

> H. I didn't expect to see any words.
> I. Was that there all the time?
> J. It wasn't that slow before.

Results for these 10 *S*s were analyzed as a second subliminal group on the assumption that they were closer to threshold (or even partially aware of the stimulus) than the .02 sec. subliminal group; the remaining eight *S*s were regarded as a "supraliminal" group. Comparing these two groups with the two groups from the previous experiment might enable one to determine the relationship between closeness to threshold, or degree of awareness, and effect of suggestion. It was assumed that the four groups lie on a continuum of awareness in the following order: Control group, Subliminal .02 group, Subliminal .5 group, Supraliminal group.

The variability introduced by the last two groups made a parametric analysis inappropriate even for log converted data (Bartlett's test yielded a chi square of 75.95). The statistical method used was the Kruskal Wallis One-Way Analysis of Variance by Ranks (Siegel, 1956). The percentage increases and decreases of the 54 *S*s in all four groups were ranked from the greatest decrease to the greatest increase. The mean ranks in each of the four groups on the percentage change from Conditions I to II and the percentage change from Conditions II to III can be seen in Table 4. Large percentage drops in productivity lower these means while large percentage increases raise them. The median percentage drop or increase in productivity is also indicated in this table.

The value of H, after correction for ties, for the change from Condition I to II was 2.61, which was not statistically significant ($df = 3$). The value of H for the change from Condition II to III was 13.33, which was significant below the .01 level. Since the mean ranks of the four groups were significantly different on the percentage change from Conditions II to III individual tests of significance were performed on differences among the mean ranks.[3] The z values resulting from these tests can be seen in Table 5. Both the .02 and .5 sec. Subliminal groups show a significantly greater drop in productivity on Condition III than the Con-

[3] For the difference between the mean ranks of any two groups, a and b,

$$z = \frac{\bar{R}_a - \bar{R}_b}{\sqrt{\sigma R^2\left(\frac{1}{n_a} + \frac{1}{n_b}\right)}}$$

where

$$\sigma R^2 = \frac{N(N+1)(N-1) - \Sigma T}{12(N-1)}$$

In this application, $N = 54$, and ΣT was 126. (It can be shown that if there were only two groups, $z^2 = H$). To test these zs for significance, the Tukey Studentized Range Test was again used, this time with $df = \infty$.

TABLE 4
MEDIAN PERCENTAGE CHANGES AND MEAN RANK CHANGES IN FOUR GROUPS

Group	N	Condition I to Condition II (WRITE MORE)		Condition II to Condition III (DON'T WRITE)	
		Med. Change (%)	M Rank	Med. Change (%)	M Rank
Control	18	+7	22.28	+15	37.42
.02 Subliminal	18	+20	28.89	−8	22.94
.5 Subliminal	10	+40	34.80	−12	16.35
Supraliminal	8	+21	27.00	+17	29.13
H			2.61		13.33
p			*NS*		$<.01$

TABLE 5
TESTS OF SIGNIFICANCE (z) BETWEEN MEAN RANKS OF FOUR GROUPS

Group	.02 Subliminal	.5 Subliminal	Supra-liminal
Control	2.76*	3.40**	1.24
.02 Subliminal		1.06	.93
.5 Subliminal			1.71

* Significant below the .05 level.
** Significant below the .01 level.

trol group, but the difference between the Supraliminal group and the Control group is not significant. Although the .5 Subliminal group drops more in productivity than the .02 Subliminal group, the difference between these groups is not significant.

It would thus appear that the suggestion effect only appeared when the suggestion stimulus was subliminal. The .5 Subliminal group merely furnishes a replication of the results in the .02 subliminal group. Why was the supraliminal suggestion ineffective? A closer look at the Supraliminal group indicates that the suggestion did affect at least one of the *S*s in this group. Only three of the eight *S*s decreased their productivity on Condition III, but one of these *S*s complied in full with the suggestion by not writing at all during this condition! Her decrease was therefore 100% from the previous trial. A second *S* decreased 56%. These two *S*s showed more response to the suggestion than any of the other *S*s in the experiment. However, their response is masked in the group data by four other *S*s who showed large increases in productivity, or negative responses to suggestion. The *S* who fully complied with the suggestion said: "I know it said 'don't write,' so I tried to concentrate on the following ones, in case it said 'write.' I couldn't see any reason for the words being there unless they were directed for me."

The *S* who decreased by 56% felt that the words were there to influence her because she had read about "subliminal advertising" but she wasn't aware that her productivity had been affected by the suggestion.

The *S*s who were contrasuggestive, i.e., increasing instead of decreasing their productivity, found some interpretation for the stimulus which explained their disregarding it.

One *S* said: "I didn't know whether it meant stop writing or not so I just went on writing what I saw."

Another *S* said: "I was just surprised to see it there. I didn't notice it before. I was trying to look at the pictures and didn't pay too much attention to it. I thought it meant don't write at that time, not to write until the light went off."

Another *S* said: "My first impulse was that it was saying 'Don't write,' but due to your previous instructions, I went ahead. Then I thought it had to do with the picture; the man leaving the woman and saying 'Don't write.' "

A final question which the experimenter tried to answer was the question of whether the individual differences in response to the suggestion stimulus were related to personality differences in *S*s. A previous study by Zuckerman and Grosz (1958) demonstrated some relationship between two of the needs measured by the Edwards Personal Preference Schedule (EPPS) (Edwards, 1954) and response to the sway test (a measure of suggestibility). The EPPS was administered to eight *S*s from the two subliminal groups who had responded in a contrasuggestive direction on Condition III (i.e., increasing rather than decreasing their productivity), and to the eight *S*s from these groups who had shown the largest positive response to the suggestion (i.e., the largest percentage decreases in productivity). Comparison of these two groups on the 15 Murray needs measured by the test yielded only one significant difference. The contrasuggestive group was significantly higher on need Affiliation (p less than .01). Since one difference out of 15 comparisons may be due

to chance, it may be concluded that no personality differences between these groups were revealed by the instrument used.

DISCUSSION

The data indicate that one of the two subliminal suggestions influenced the productivity of the *S*s in describing the pictures. It is not clear why the suggestion "Don't Write" was more effective than the suggestion, "Write More." The suggestion Don't Write seemed to have the effect of counteracting a practice effect, for the control group steadily increased their productivity going from Condition I to Condition III. One can speculate that there were two opposing motivations in the last condition: (*a*) the usual positive motivations of experimental *S*s for personal achievement or to please the experimenter; (*b*) a fatigue effect toward the end of the experiment resulting in a disposition to exert less effort. The suggestion Don't Write may have reinforced the increasing disposition to slow down. More work is needed to explore the relationship between subliminal suggestion and the intensities and types of needs or motives being reinforced.

The degree of "awareness" or recognition plays a crucial role in disputes over the phenomena of subliminal perception. Goldiamond (1958) feels that the phenomena can be explained in terms of "partial awareness" which the *S* cannot communicate because of the lack of quantitative specificity in the verbal indicators used to define "perception." This question can be treated in empirical terms by asking "what is the relationship between the subliminal suggestion effect and the 'degree of awareness' of the subliminal stimulus?" This question has been investigated in this experiment by: (*a*) examining the relationship between the magnitude of response to suggestion and the threshold for the suggestion stimulus and (*b*) comparing two subliminal groups (differing markedly in the time of exposure of the subliminal stimulus), a supraliminal group and a control group. There is some tendency for a greater suggestion effect in the *S*s operating below, but closer to threshold. However, neither the correlation between threshold and magnitude of response, nor the difference between the .02 sec. and .5 sec. Subliminal groups reached significance. Both of the subliminal groups did differ significantly from the control group, whereas the supraliminal group did not. *S*s who became completely aware of the suggestion seemed to respond in a dichotomous fashion: either conforming strongly with the suggestion or responding strongly in a contrasuggestive direction. The latter tendency was actually more prevalent. This points up the difference between the present experiment and those in which the *S* is given a discrimination task. When the *S* is looking for an answer to a problem, he will, of course, use the answer that the experimenter gives him, and the greater the intensity of the cue the more likely he will respond. However, when the subliminal cue is not directly relevant to the primary task on which the *S* is focusing, he is quite likely to reject or ignore it as an annoying distraction. This is what many of the supraliminal *S*s in this experiment apparently did. The "distracting" or "irrelevant" nature of the subliminal cue may account for the difficulty which the majority of *S*s had in seeing it even though it was exposed for a relatively long interval following the picture to which they were attending.

SUMMARY

Two experiments were performed to study the effect of subliminal and supraliminal suggestion on productivity, or number of words written in describing a picture. An experimental group was given a task of describing in writing a TAT picture presented tachistoscopically 10 times at increasing exposure levels. In the second condition a subliminal stimulus, the words WRITE MORE overlapped with the TAT picture for .02 seconds. In the third condition the subliminal suggestion was DON'T WRITE. A different TAT card was used for each condition and the cards were rotated among conditions in three orders so that each card was presented in each condition an equal number of times. The control group performed the same task on the three cards but instead of a subliminal stimulus only a blank flash occurred in Conditions II and III. There were 18 *S*s each in the experimental group and control groups. Significant differences between groups as a function of conditions were found. Whereas the control group increased their

writing rate steadily from Condition I to Condition III, the experimental group decreased their rate on Condition III.

A second experiment was performed to see the effect on productivity of the same suggestions at supraliminal levels. The exposure intervals of the suggestions were increased to .5 sec. Ten of the group of 18 tested under these conditions still could not report awareness of the suggestions, although some of them had "partial awareness." These *S*s showed the same effect as *S*s in the .02 sec. condition. The eight *S*s who recognized at least one of the suggestions did not show a consistent response to them. Some responded in the suggested manner, i.e., markedly increasing or decreasing their rate on request, while others ignored the suggestion or did not interpret it as "instructions." The net effect was closer to the response of the control group than to that in either of the subliminal groups. An hypothesis formed from these data is that subliminal suggestion may produce some effect in the region just below threshold. When the suggestion becomes supraliminal its distracting effect causes a contrasuggestive response in some *S*s. No marked personality differences (as measured by the EPPS) were found between suggestive and contrasuggestive *S*s in the two subliminal groups.

REFERENCES

ADAMS, J. K. Laboratory studies of behavior without awareness. *Psychol. Bull.*, 1957, **54**, 383–405.

EDWARDS, A. L. *Experimental design in psychological research*. New York: Rinehart, 1950.

EDWARDS, A. L. *Edwards Personal Preference Schedule. Manual*. New York: Psychological Corp., 1954.

ERIKSEN, C. W. Unconscious processes. In M. R. Jones (Ed.), *Nebraska symposium on motivation*. Lincoln: Univer. Nebraska Press, 1958. Pp. 169–227.

GOLDIAMOND, I. Indicators of perception: I. Subliminal perception, subception, unconscious perception: An analysis in terms of psychophysical indicator methodology. *Psychol. Bull.*, 1958, **55**, 373–411.

MCCONNELL, J. V., CUTTER, R. L., & MCNEIL, E. B. Subliminal stimulation: An overview. *Amer. Psychologist*, 1958, **13**, 229–242.

SCHEFFE, H. A method for judging all contrasts in the analysis of variance. *Biometrika*, 1953, **40**, 87–104.

SIEGEL, S. *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill, 1956.

ZUCKERMAN, M., & GROSZ, H. J. Suggestibility and dependency. *J. consult. Psychol.*, 1958, **22**, 328.

(Received April 28, 1959)

Journal of Abnormal and Social Psychology
Vol. 60, No. 3, 1960

CONFUSION OF FIGURATIVE AND LITERAL USAGES OF WORDS BY SCHIZOPHRENICS AND BRAIN DAMAGED PATIENTS[1]

LOREN J. CHAPMAN[2]
University of Chicago

MANY clinical writers have commented on an aspect of the thinking disorder of schizophrenic and brain damaged patients which they describe as a confusion of literal usages of words with more figurative or metaphoric usages. The present study is an attempt to investigate such phenomena by means of objective measures.

Clinical writers have seldom defined the terms literal, figurative, and metaphoric, but have instead relied on the use of example to indicate their meaning. However, linguists have offered explicit definitions which should be of use to clinicians who are concerned with this area.

A figure of speech is usually defined by linguists as an extension of the applicability of a word. Figures of speech are often divided into such categories as metaphor, metonomy, and synecdoche (Sturtevant, 1947, pp. 137–139). Metaphor is usually defined as the extension of the applicability of a word in order to suggest a similarity or analogy between two objects, ideas, or events, e.g., "pull strings" for "use influence." Such metaphoric extensions may be: (*a*) from one sense modality to another, e.g., *loud* color; (*b*) from one object to another within the same sense modality, e.g., the use of the term *moon* to describe man-made satellites; (*c*) from a sense modality to a notion that ordinarily is not susceptible to direct sensory apprehension, e.g., *warm* feelings. Freshly coined and original metaphors, which include those of poetry and other similar literature, are readily recognized as such and usually require effort on the part of the listener to grasp their meaning. With continued usage, a metaphor becomes incorporated into the language to the point that it is seldom recognized as a metaphor, e.g., *high* pitch.

The present investigation is concerned with metaphors which are fairly well incorporated into everyday usage but are still recognizable as metaphors. It is primarily concerned with the third kind of shift listed here, shifts from a sense modality to the description of subjective judgments, emotional states, or the quality of interpersonal interactions.

Clinical writers have described the thinking disorder in terms of a disturbance in the use of this type of metaphor both in schizophrenics (Arieti, 1955; Bateson, 1956; Bleuler, 1950; Goldstein & Scheerer, 1944; Muncie, 1937; Storch, 1924) and organics (Goldstein, 1948; Head, 1926; Jakobson & Halle, 1956). All of these writers attribute both inappropriately literal and inappropriately metaphoric utterances to patients. However, most of these writers (except Bateson and Muncie) state that the apparent metaphoric utterances are seldom or never truly metaphoric, but are, instead, meant in a literal way by the patient. For example, Storch (1924, pp. 5–9) construed a schizophrenic's reference to a person's train of thought as a "thread" and his will as a "spark" as referring to sensory images of a physical spark and a physical thread without any metaphoric intent on the patient's part.

Such clinical views seem to imply that if the patients were presented with words in context and asked to indicate their meaning, they would often misinterpret figurative usages as if they were literal but would not commit the reverse error of misinterpreting literal usages as if they were figurative. That is, such a patient might make a literal misinterpretation of the statement, "The only way Tom can get appointed to West Point is to pull strings," as meaning, "It is necessary for him to grab a cord in each hand and yank it"; however, he would not commit the converse *figurative* misinterpretation of "pull strings" in the statement, "The

[1] This study was supported by a research grant (M-1416) from the National Institute of Mental Health of the National Institutes of Health, United States Public Health Service.

The author is indebted to Alvin Burstein for his valuable suggestions in the preparation of the manuscript, to Dorothy Day for assistance in the collection and analysis of the data, and to Catherine Becker for assistance in analysis of the data.

[2] Now at the University of Kentucky.

only way for Jim to lace his boots is to pull strings." Nevertheless, Bateson (1956) and Muncie (1937) specifically state that schizophrenics commit this latter kind of error. Muncie believes that it is more prominent in schizophrenia than excessive literalness but that both kinds of misinterpretation occur. For Bateson, the tendency of schizophrenics to make both kinds of misinterpretation is a central feature of his theory of schizophrenia.

In the present investigation, schizophrenics, brain damaged patients, and normal *S*s were compared on these two kinds of misinterpretation.

PROCEDURE

Instrument

A multiple-choice paper-and-pencil test was constructed. The following examples are representative of the items used in the test:

9. David turned yellow when he faced the enemy. This means:
   - A. David became cowardly.
   - B. David became hungry.
   - C. David's skin became discolored.

31. Miss Bailey's illness turned her yellow. This means:
   - A. Miss Bailey became cowardly.
   - B. Miss Bailey became hungry.
   - C. Miss Bailey's skin became discolored.

In Item 9 above, the correct answer is, of course, "David became cowardly." Choosing "David's skin became discolored" would reflect an inappropriately literal interpretation of the word "yellow." In Item 31, the reverse is true; "Miss Bailey's skin became discolored" is the correct answer, and "Miss Bailey became cowardly" is inappropriately figurative.

The third alternative, which interprets yellow to mean hungry, is neither correct nor a predicted error due to misinterpretation. But inclusion of this alternative has a special use in the design. Errors are made for many reasons besides the kind of misinterpretation under investigation, and the bases of these errors are mostly unidentified. For example, many schizophrenics are noncooperative and respond carelessly, mark randomly, mark by position, or mark in response to other characteristics of the test which are difficult for the examiner to identify. If such unidentified errors are equally distributed among the three alternatives, then a score of predicted error markings minus "unidentified" error markings should be a relatively pure measure of the predicted error tendency.

Another way of putting it is that there should be, by chance, as many errors made for unidentified reasons hidden among the predicted misinterpretation markings as there are markings of these "unidentified" error alternatives themselves. Therefore, we may obtain for each *S* a relatively pure measure of literal misinterpretation if we subtract from the number of his literal error markings the number of his "unidentified" error markings. This will be called the "corrected literal score." The "corrected figurative score" is obtained by following the same principle.

The instrument was a 40-item paper-and-pencil test[3] which consisted of 20 items which permit literal misinterpretations and 20 items which permit figurative misinterpretations. The items were constructed in pairs, as in the example above, with one item testing for a literal misinterpretation of a given word or phrase and the other for a figurative misinterpretation of that same word or phrase. In each case, the "correct" interpretation was determined by the context. The 20 figurative and the 20 literal items were intermixed in random order, with the restriction that the two items testing the literal and figurative aspects of the same phrase were separated by several other items. The order of the three kinds of alternatives was randomized among items with the restriction that the two items within each pair had the same order.

Subjects

Three groups of *S*s were tested[4]: 36 normal *S*s, 41 brain damaged *S*s, and 51 schizophrenics. The normal *S*s consisted of 23 medical and surgical patients at the West Side VA Hospital, Chicago, Illinois, and 13 Chicago firemen. No normal *S* was used for whom there was any indication either of psychiatric or neurological disorder.

The schizophrenic group, patients at Chicago State Hospital, had all received a hospital diagnosis of chronic schizophrenia, with their first hospitalization for psychosis occurring at least five years previously and with continuous hospitalization for at least three months previous to testing. Almost all of the schizophrenic patients had spent most of the last 10 years in the hospital. Other prerequisites for selection were absence of any indication of either cerebral organic involvement or mental deficiency, no current drug therapy, and no electric shock therapy within the past three months. No selection was attempted by diagnostic subtype.

The brain damaged patients consisted of 29 patients at Hines VA Hospital and 12 patients at Chicago State Hospital. The group included 13 diagnoses of cerebral

[3] Copies of the test may be obtained from the author, or may be ordered from the American Documentation Institute. Order Document No. 6237, from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting $1.25 for microfilm or $1.25 for photocopies. Make checks payable to Chief, Photoduplication, Service, Library of Congress.

[4] The author is indebted to Kalman Gyarfus and J. Herbert Maltz of the Chicago State Hospital for permission to test schizophrenic patients for this experiment, to Gabrielle P. Ferrazzano for permission to test patients at the USPHS Hospital, to Roy Brener for assistance in obtaining patients at Hines Hospital, to Meyer Williams for assistance in obtaining *S*s at the West Side VA Hospital, and to Robert J. Quinn for his help in obtaining the cooperation of Chicago firemen.

vascular disorder, 6 of alcoholism, 6 of multiple sclerosis with cerebral disorder, 3 of brain tumor, 2 of head injury, 1 of Huntington's Chorea, 1 of syphilitic paresis, and 9 of unspecified cerebral damage. The organics, as a group, had been hospitalized for a shorter term than the schizophrenics. About one-third of the organics had entered during the previous three months, and none had been hospitalized longer than five years.

Since these patients were being tested in order to study the nature of cognitive disturbance when it results from brain injury, only those patients were included who appeared by clinical interview to have some intellectual loss or disturbance. This decision was usually made on the basis of the patient's difficulty in communication, serious disturbance of memory, or severe lack of alertness as observed either in a clinical interview or in his performance on the CVS, a brief intelligence scale (Hunt, French, Klebanoff, Mensh, & Williams, 1948). No patient was included in the brain damaged group who appeared also to be schizophrenic.

The mean age was 43.4 years for the schizophrenics, 44.7 for the brain damaged, and 43.1 for the normals, while mean years of schooling was 10.2 for the schizophrenics, 10.2 for the brain damaged, and 10.6 for the normals. Although the groups did not differ significantly with respect to age or education, the schizophrenic group was lower ($p < .001$) than the organics on functioning IQ as determined by the CVS , and the organic group was lower ($p < .01$) than the normal group, presumably due to the disruptive effects of the schizophrenic and organic pathologies. The mean IQ was 86 for the schizophrenics, 104 for the organics, and 113 for the normals.

The *Ss* were assumed to be of approximately equal premorbid intelligence because of the equivalent education. An additional approximation of premorbid intelligence was obtained from the item number of the last vocabulary word answered correctly. As suggested by Becker (1956), this appears to be a better estimate than the total number of vocabulary words answered correctly because one effect of psychopathology is the production of uneveness in performance. On this measure, the schizophrenics scored a mean of 23.3, the organics 26.2, and the normals 27.7. None of the differences was significant.

RESULTS

All between-group differences were evaluated by a two-tailed Mann-Whitney test and all within-group differences by a two-tailed Wilcoxon test.

Table 1 lists the median error scores for each of the three groups. Each of the three groups marked significantly more literal misinterpretations than "unidentified" error alternatives and significantly more figurative misinterpretations than "unidentified" error alternatives ($p < .001$ in each case).

TABLE 1
MEDIAN SCORES OF PREDICTED AND UNIDENTIFIED ERRORS

	Schizophrenic	Brain damaged	Normal
Figurative error items			
Figurative errors	2.6	1.9	1.0
Unidentified errors	.3	.2	.0
Corrected Figurative error score (Figurative minus Unidentified)	1.4	1.4	.9
Literal error items			
Literal errors	5.3	.6	.4
Unidentified errors	.7	.1	.0
Corrected Literal error score (Literal minus Unidentified)	3.1	.4	.4

Comparison of schizophrenic and normal performance. By corrected literal score the schizophrenics were more susceptible to literal misinterpretations than the normal *Ss* ($p < .001$). However, there was only a trend towards a difference on the corrected figurative score ($p < .08$). It would appear, therefore, that the tendency toward literal misinterpretations is greater than that toward figurative misinterpretations by schizophrenics on this test. By comparing, for the schizophrenic group alone, the corrected literal score with the corresponding corrected figurative score, it was found that the literal misinterpretations exceeded the figurative ($p < .01$). The normal *Ss* showed a trend toward the opposite error preference, but not significantly so ($p = .20$).

These findings tend to support Bateson's statements, in that both kinds of misinterpretations tend to be made more often by schizophrenics than by normal *Ss*. The fact that the normal *Ss* also chose the misinterpretations more often than the unidentified errors shows that the misinterpretations are not unique to pathological patients, but rather that their greater frequency in schizophrenics is an exacerbation of a normal error tendency. The greater prevalence of the literal misinterpretations is in keeping with the observations of Arieti (1955), Bleuler (1950), Goldstein (1944), and Storch (1924). This difference in frequency between two kinds of misinterpretation on this test for a single group of *Ss*, however, may be

a function of the particular literal and figurative items offered. Therefore, any finding concerning comparative susceptibilities of any one group to the two kinds of misinterpretations on this test may be lacking in generality. Nevertheless, any finding of a difference *between* groups on this test is not subject to this stricture because the same items were given to all *Ss*.

Comparison of brain damaged patients and normal Ss. Unlike these schizophrenics, the brain damaged patients showed no tendency toward more literal misinterpretations, as measured by the corrected error score, than the normal *Ss* ($p = .89$). They showed only a trend toward a greater number of figurative misinterpretations than normals ($p = .11$).

Comparison of the brain damaged patients' preference for one or the other kind of misinterpretation, as measured by the corrected scores, showed that they more often make figurative misinterpretations than literal misinterpretations ($p < .05$). The normal *Ss*, it will be recalled, showed no significant preference, while the schizophrenics showed the opposite preference.

Comparison of schizophrenic and organic brain damaged patients. The schizophrenics made more total errors than the brain damaged patients on both tests, but this may merely reflect differing degrees of pathology between the particular samples of patients tested. Comparisons of gross error tendency, which are so commonly reported in the literature, do not appear to be meaningful on any single test unless the two groups are equated in some way on over-all pathology—an extremely difficult task.

The schizophrenics and the organics can be meaningfully compared, however, on their relative preferences for the two kinds of misinterpretations. The schizophrenics, as mentioned above, made more literal misinterpretations than figurative, while the brain damaged patients showed the opposite preference. This difference was more directly examined by computing for each *S* the difference between his two corrected error scores. On this measure, the schizophrenics and organics differed ($p < .001$), and they were on opposite sides of normal *Ss*. The schizophrenics differed significantly from the normals on this score, ($p < .01$), but the brain damaged patients did not ($p = .26$).

Although the schizophrenics and brain damaged patients differed in the kind of misinterpretation preferred, some of the patients in each group showed a preference opposite to that of their group as a whole. For the schizophrenics, 61% preferred the literal misinterpretation, 25% preferred the figurative, and 14% showed no preference. For the brain damaged patients, the preferences were 29% for the literal misinterpretations, 54% for the figurative, and 17% with no preference. It is interesting to conjecture whether the *Ss* with atypical scores in each group may differ in other ways from the main body of patients. However, these atypical scores may have arisen merely from the unreliability of the test. For the schizophrenic patients, the Kuder-Richardson (Formula 20) reliabilities were .47 for the corrected figurative error score and .55 for the corrected literal error score. The corresponding values for the brain damaged patients were .49 and .43, respectively.

SUMMARY

A test was designed to measure the tendency to make inappropriately literal and inappropriately figurative misinterpretations of written statements. The schizophrenic patients made significantly more literal misinterpretations than figurative, while brain damaged patients showed the opposite effect.

A score of literal misinterpretation minus figurative misinterpretation distinguished the two groups and placed them on opposite sides of the normal *Ss*.

REFERENCES

ARIETI, S. *Interpretation of schizophrenia.* New York: Robert Brunner, 1955.

BATESON, G., JACKSON, D. N., HALEY, J., & WEAKLAND, J. Toward a theory of schizophrenia. *Behav. Sci.*, 1956, **1**, 251–264.

BECKER, W. C. A genetic approach to the interpretation and evaluation of the process reactive distinction in schizophrenia. *J. abnorm. soc. Psychol.*, 1956, **53**, 229–236.

BLEULER, E. *Dementia praecox or the group of schizophrenias.* New York: Int. Univer. Press, 1950.

GOLDSTEIN, K. *Language and language disturbances.* New York: Grune & Stratton, 1948.

GOLDSTEIN, K., Methodological approach to the study of schizophrenic thought disorder. In J. S. Kasanin (Ed.), *Language and thought in schizophrenia.* Berkeley: Univer. California Press, 1944.

HEAD, H. *Aphasia and kindred disorders of speech.* Vol. 1. New York: MacMillan, 1926.

HUNT, W. A., FRENCH, ELIZABETH G., KLEBANOFF, S. G., MENSH, I. N., & WILLIAMS, M. Clinical possibilities of an abbreviated individual intelligence scale. *J. consult. Psychol.*, 1948, **12**, 171–173.

JAKOBSON, R., & HALLE, M. *Fundamentals of language.* The Hague: Mouton, 1956.

MUNCIE, W. The psychopathology of metaphor. *Arch. Neur. Psychiat., Chicago*, 1937, **37**, 796–804.

STORCH, A. The primitive archaic forms of inner experiences and thought in schizophrenia. *Nerv. ment. Dis. Mongr. Ser.*, 1924, No. 36.

STURTEVANT, E. H. *An introduction to linguistic science.* New Haven: Yale Univer. Press, 1947.

(Received July 20, 1959)

Journal of Abnormal and Social Psychology
Vol. 60, No. 3, 1960

LEARNING A SOCIAL STRUCTURE[1]

CLINTON B. DE SOTO
Johns Hopkins University

For some years social psychologists have shown increasing interest in mathematical properties of social structures, ranging from investigations of the extent to which sociometric choices are mutual to experimental manipulations of the centrality of communication networks. This interest has been accompanied, however, by a curious lack of concern with the subjects' (*S*s') views of these properties. For example, in the studies of mutuality of sociometric choices, the *S*s are not ordinarily asked if *they* expect the choices to be mutual. Only recently, in such work as Tagiuri's (1958) generalization of sociometry, relational analysis, and Heider's (1958) discussion of naive psychology, has there been serious concern with the *S*s' perception of such properties.

Apparently psychologists have assumed that *S*s cannot or do not appreciate these properties. Yet folklore is full of essentially mathematical statements about interpersonal relations: "The way to get someone to like you is to like him," "You can't serve two masters," "There's always a faster gun somewhere." And recent questionnaire studies (De Soto & Kuethe, 1958, 1959) have demonstrated that *S*s ascribe such properties rather freely to interpersonal relations, sometimes accurately, sometimes inaccurately.

Historically social psychologists have repeatedly been compelled to recognize the importance of cognitive processes in man, to the need, for example, for crediting people with role expectations as well as role behaviors. This fact makes it appear unwise to ignore their readiness to assign mathematical properties to interpersonal relations. It makes it important to look for effects of such attributions of properties in activities besides answering questionnaries. There are in the literature many findings which are suggestive of such effects. For example, Kogan and Tagiuri (1958) found that people expect those they choose to choose one another, thus evidencing, for a special case, their expectation that sociometric choices will be mutual.

One task in which *S*s' tendencies to assign properties to interpersonal relations might be expected to have especially prominent effects is that of learning a social structure. Learning a social structure means learning the relations among a set of people. It is a task that usually faces a person when he enters a social group, and it can be quite difficult as indexed by the information he must assimilate. It is a reasonable prediction that his assignments of properties to the relations which form the structure, operating as expectancies or hypotheses, will facilitate or hinder his learning of the structure according to their validity. The present study was designed to verify this general prediction and also to examine in more detail some of the psychological processes involved.

The study was basically a verbal learning experiment in which *S*s learned miniature social structures. The dependent variable was the number of trials required. There were two independent variables: the interpersonal relation that formed the structure, and the mathematical properties of the structure. In their questionnaire study of the properties assigned to various interpersonal relations, De Soto and Kuethe (1959) found that the relations fell into two main classes: those regarded as asymmetric, which were also regarded as transitive, and those regarded as symmetric, some of which were also regarded as transitive and some of which were not. For the present study, the relation "influences" was chosen to exemplify the relations people expect to be asymmetric and transitive, "likes" was chosen to exemplify those expected to be symmetric and transitive, and "confides in" was chosen to exemplify those expected to be symmetric but not transitive.

Structures characterized by eight different sets of mathematical properties were used. Four-man structures were used because they were the largest *S*s could learn with relative

[1] This work was supported in part by Grant NSF-G4827 from the National Science Foundation. The writer is grateful to Martin N. Narun and Daniel M Glasner for aid in the collection of data.

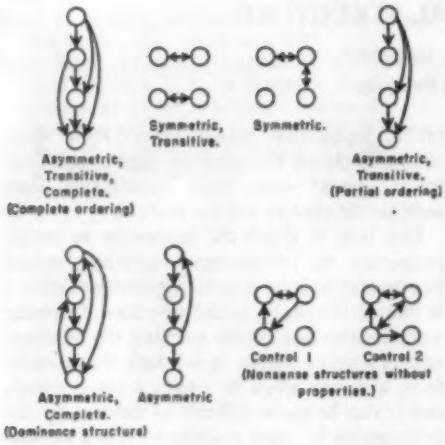

FIG. 1. Structures used in the learning experiment.

ease in pilot studies. The structures are depicted in Fig. 1, with circles representing people and arrows a relation. An arrow going from one circle to another means the first person stands in the given relation to the second; if the relation is "influences," it means the first person influences the second. The requirements for the various properties may be defined informally as follows: For symmetry, the arrows must go in both directions, or not at all, between a pair of individuals. For asymmetry, the arrows must never go in both directions between a pair of individuals. The requirement for transitivity is more complex. If there is an arrow from Individual *a* to Individual *b*, and another from Individual *b* to Individual *c*, then, for the transitive property to hold, there must be one from *a* to *c*. For the complete property to hold, there must be at least one arrow connecting the individuals in all possible pairs. More formal definitions of these and other properties are given by Coombs, Raiffa, and Thrall (1954) and Kershner and Wilcox (1950). The names given to some of the structures in Fig. 1 are based on their properties.

The task given each *S* in the experiment was to learn one of the structures. He did not see a drawing of the structure, however, nor was the relation represented by abstract arrows. Instead, the arrows were replaced by one of the flesh-and-blood relations, "influences," "likes," or "confides in," and he had to learn, by repeated trial and error, whether or not Jim likes Ray, whether or not Ray likes Stan, and so on for the 12 ordered pairs of names which exist in a four-man structure. This arrangement of the task avoided unintended suggestions to the *S* about properties of the structure and approximated the real-life case where people learn a social structure by observing the pair-wise interactions among its members rather than from devices like organizational charts.

An experimental condition in this study was a combination of one of the abstract structures and one of the interpersonal relations. There were 19 experimental conditions. Each of the first three structures in Fig. 1 was formed from each of the three relations, producing 9 conditions, and the remaining five structures were formed from the relations "influences" and "likes," producing 10 more conditions. The first 9 conditions were the basic experimental ones, since a crucial interaction was predicted for them. On the basis of the conclusion by De Soto and Kuethe (1959) that college students (used as *S*s in that study and this) are alert to power based orderings, it was predicted that they would expect the relation "influences" to be asymmetric and transitive and, possibly, complete and would therefore learn a complete ordering based on "influences" more readily than one of the symmetric structures, while for "likes" and "confides in," which *S*s expect to be symmetric, the converse would be true. The remaining conditions were included to provide further analytical possibilities.

METHOD

The experimental method resembled that of traditional paired-associates learning experiments. The *S*s, male undergraduates at Johns Hopkins University, were assigned randomly, nine to each of the 19 experimental conditions, and were run individually. Their task is described well by the instructions given them:

> This experiment is on the learning of interpersonal relationships. There are four people involved, Bill, Jim, Ray, and Stan, and your job is to learn who likes who among these people.
>
> On the upper side of each of these cards are printed the names of two of the four people. In each case, the first person either does or does not like the second. If he does, it says so on the other side of the

card. If he doesn't, nothing is printed on the other side of the card.

When I tell you to begin, look at the names on the top card. If you think the first person likes the second, say so. If you don't think he does, say so. For example, if the first name were Sam and the second were Charlie, you would say either "Sam likes Charlie" or "Sam doesn't like Charlie." Then, turn the card over and see whether or not he does, and place the card on a new pile.

Go through the rest of the cards the same way, and then I will shuffle them and you can go through them again. As we go through the cards again and again, you should learn better and better who likes who.

Take your time. Don't rush.

The above are the instructions given to all *S*s who learned a social structure based on "likes." The instructions given to *S*s who learned a social structure based on "influences" or "confides in" were identical in form except that the appropriate one of the following definitions was added: "To influence someone means to have some degree of control over what he does"; "To confide in someone means to have complete faith and confidence in him so that you trust him with secrets or rely on him for help."

Each *S* worked with a deck of 12 cards appropriate for his experimental condition.

RESULTS

The difficulty of learning the various social structures was measured by the mean number of trials required by the *S*s to learn them to a criterion of two successive correct trials. These means are shown in Table 1. The standard errors of the means, highly correlated with their values, range from 1.06 to 4.27.

The hypothesized interaction between relations and structures is evident in Table 1. The *S*s learned a complete ordering based on "influences" very readily, but had somewhat more difficulty with symmetric structures based on "influences." They had difficulty learning a complete ordering based on "likes" or "confides in," but learned symmetric structures based on these relations easily. An analysis of variance performed on the data for these nine conditions yielded an interaction mean square significant at the .01 level.

TABLE 1
MEAN TRIALS TO CRITERION

Structure	Relation		
	Influences	Likes	Confides in
Asymmetric, transitive, complete (complete ordering)	8.9	14.7	16.2
Symmetric, transitive	12.9	10.9	9.1
Symmetric	11.7	13.0	10.9
Asymmetric, transitive (partial ordering)	12.1	14.7	
Asymmetric, complete (dominance structure)	26.4	23.7	
Asymmetric	16.3	29.1	
Control 1	16.1	20.6	
Control 2	16.7	21.0	

Examination of the remaining ten means shows that the partial ordering was learned relatively easily, at least when based on "influences," but the other conditions were all difficult, some of them exceedingly so. Particularly striking is the finding that the asymmetric and complete structure based on "influences" took substantially longer to learn than any of the nonsense structures, rather than showing facilitation from the fact that its properties were partly as expected by the *S*s.

DISCUSSION

The experimental results clearly support the general hypothesis that people's assignments of mathematical properties to an interpersonal relation influence the ease with which they learn a social structure based on the relation, and they invite analysis of the psychological processes involved.

Perhaps the most remarkable finding is this: although the combination of the asymmetric, transitive, and complete properties (complete ordering) greatly facilitates the learning of a social structure based on "influences," the combination of only the asymmetric and complete properties (dominance structure) seems to make the learning task even more difficult than it is for structures without properties. The *S*s failed completely to profit from the latter combination of properties, although in principle it too could be a valuable aid to learning a structure. Why?

It appears an inescapable conclusion that for the *S*s these properties were not independent of one another, even though they are independent mathematically, and even though *S*s are able to ascribe them separately to relations when questioned (De Soto & Kuethe, 1958, 1959). De Soto and Kuethe noted a tendency for *S*s usually to ascribe the transitive property whenever they ascribed the asymmetric property to a relation. Evidently this tendency to couple the properties is so

powerful that it prevents people from using the properties in isolation when learning a social structure. People expect mathematical properties in social structures, to be sure, but they expect them in bundles, and incomplete bundles are of no use to them.

It is as if the *S*s had a theory about the social structure, a mathematical model for it which was very helpful when its properties corresponded with those of the social structure, but which had to be applied in an all-or-nothing manner in the sense that it was not readily abridged in the face of some lack of correspondence and became instead a hindrance. Such a theory, marked by certain essential properties but doubtlessly skeletal and sketchy in other respects, seems most aptly called a *schema*, roughly following Bartlet (1932) and Woodworth (1938).

The present data, then, might be interpreted to indicate that the *S*s possessed an ordering schema which they automatically applied to a social structure based on the relation "influences," although not to one based on "likes" or "confides in." "Influences" might be said to tap or arouse people's ordering schema. This ordering schema, the residue perhaps of countless experiences with social and nonsocial orderings in the world, seems to be marked by the properties of asymmetry, transitivity, and, possibly, completeness. Thus the *S*s expected these properties in a social structure based on "influences," but not as independent expectations; rather, as derivatives of the overriding expectation that the structure was an ordering. Perhaps the dominance structure was more difficult than the nonsense structures because it perpetually titillated and tantalized the *S*s' ordering schema, preventing the suppression of it that occurred for the nonsense structures.

Evidence that more sophisticated students of social structures also possess and can be victims of an ordering schema is found in many writings about peck rights among birds. Birds establish peck rights on the basis of pair-wise encounters, a process which insures that the eventual social structure is asymmetric and complete, but not that it is transitive. The birds, being simple-minded creatures not possessed of an ordering schema, thus innocently build a social structure that puzzles and provokes more thoughtful creatures who write about the "pecking out of order," "curious exceptions," "irregularities," "imperfect orderings," "lack of a hierarchy," "polygonal dominance," "cyclic pecking," "triangular pecking," and "complex, triangular hierarchies" (Allee, 1931; Collias, 1944; Murchison, 1935; Murphy, Murphy, & Newcomb, 1937; Schjelderup-Ebbe, 1935). People consider this primitive and simple social structure surprising and complex because they expect an ordering. To turn the thing around, it may be that the expectation of an ordering tends to *produce* an ordering, that human social groups would show fewer orderings and more dominance structures and even property free structures if it were not for the shared expectation of orderings.

The finding that the partial ordering was learned quite readily when based on "influences," with the *S*s' ordering schema presumably aroused, suggests that the property of completeness is a less integral part of the ordering schema than the properties of asymmetry and transitivity. However, any partial ordering can be mathematically treated as made up of two or more complete orderings (Coombs & Kao, 1955; Kershner & Wilcox, 1950), and it is possible that the *S*s accomplished essentially this analysis of the partial ordering.

The property of symmetry seems to be fairly easy to detect and utilize even when it is not expected, as in the structures based on "influences." The results suggest that for "likes" and "confides in" a symmetric and transitive structure is learned easier than a merely symmetric one, but the differences are not large, and it should be noted that the so-called symmetric and transitive structure of Fig. 1 is only *vacuously* transitive (Kershner & Wilcox, 1950), having too few members to be truly transitive. Thus, it is impossible to say more than that a lack of transitivity is not the severe handicap in learning symmetric structures that it is in learning asymmetric ones. This conclusion is consistent with the view expressed earlier (De Soto & Kuethe, 1959) that the primary property ascribed to these relations is symmetry, and that transitivity is ascribed secondarily, only after long experience which teaches that relations

like "likes" are commonly transitive, dividing people into cliques or friendship groups. The grouping schema, as applied to interpersonal relations at least, is a higher development of the symmetry schema.

SUMMARY

By means of a learning experiment it was shown that a social structure is easiest to learn when it possesses the mathematical properties people expect in it. People attribute the properties of asymmetry, transitivity, and completeness (the properties of an ordering) to the relation "influences." If a social structure formed of the relation "influences" has these properties, they learn it readily. If it is symmetric, or if it has no mathematical properties, they have more difficulty; and if it is asymmetric and complete but not transitive (a dominance structure), they have great difficulty. Similarly, structures based on "likes" or "confides in" are learned most readily when they are symmetric, as people expect them to be. The findings were interpreted as reflecting the operation of schemas in the *S*s.

REFERENCES

ALLEE, W. C. *Animal aggregations: A study in general sociology.* Chicago: Univer. Chicago Press, 1931.

BARTLETT, F. C. *Remembering.* Cambridge: Cambridge Univer. Press, 1932.

COLLIAS, N. E. Aggressive behavior among vertebrates. *Physiol. Zool.*, 1944, **17**, 83–123.

COOMBS, C. H., & KAO, R. C. Non-metric factor analysis. *Engng. res. Bull.* No. 38. Ann Arbor: Univer. Michigan Press, 1955.

COOMBS, C. H., RAIFFA, H., & THRALL, R. M. Some views on mathematical models and measurement theory. *Psychol. Rev.*, 1954, **61**, 132–144.

DESOTO, C. B., & KUETHE, J. L. Perception of mathematical properties of interpersonal relations. *Percept. mot. Skills*, 1958, **8**, 279–286.

DESOTO, C. B., & KUETHE, J. L. Subjective probabilities of interpersonal relationships. *J. abnorm. soc. Psychol.*, 1959, **59**, 290–294.

HEIDER, F. *The psychology of interpersonal relations.* New York: Wiley, 1958.

KERSHNER, R. B., & WILCOX, L. R. *The anatomy of mathematics.* New York: Ronald, 1950.

KOGAN, N., & TAGIURI, R. Interpersonal preference and cognitive organization. *J. abnorm. soc. Psychol.*, 1958, **56**, 113–116.

MURCHISON, C. The experimental measurement of a social hierarchy in *gallus domesticus:* I. The direct identification and direct measurement of social reflex no. 1 and social reflex no. 2. *J. gen. Psychol.*, 1935, **12**, 3–39.

MURPHY, G., MURPHY, L. B., & NEWCOMB, T. M. *Experimental social psychology.* (rev. ed.) New York: Harper, 1937.

SCHJELDERUP-EBBE, T. Social life of birds. In C. Murchison (Ed.), *Handbook of social psychology.* Worcester, Mass.: Clark Univer. Press, 1935.

TAGIURI, R. Social preference and its perception. In R. Tagiuri & L. Petrullo (Eds.), *Person perception and interpersonal behavior.* Stanford, Cal.: Stanford Univer. Press, 1958.

WOODWORTH, R. S. *Experimental psychology.* New York: Holt, 1938.

(Early publication received October 22, 1959)

Journal of Abnormal and Social Psychology
Vol. 60, No. 3, 1960

REPEATED FRUSTRATIONS AND EXPECTATIONS IN HOSTILITY AROUSAL[1]

LEONARD BERKOWITZ
University of Wisconsin

IN THEIR now classic monograph, *Frustration and Aggression*, Dollard, Doob, Miller, Mowrer, and Sears (1939) hypothesized that the total strength of the "instigation to aggression" within an individual is a positive function of three aspects of the frustrations he had suffered: "1) the strength of the instigation to the frustrated response, 2) the degree of interference with the frustrated response, and 3) the number of frustrated response-sequences" (p. 28). Even accepting Miller's later admonition that these propositions refer to the *instigation* to aggression (for simplicity, we might say "aggressive drive") and not aggressive *acts* (1941), many critics still objected to their failure to differentiate among qualitatively different classes of frustrations. For example, Maslow (1941) maintained only threatening frustrations give rise to aggression, while mere deprivations presumably do not. In the same vein, Pastore (1952) proposed that arbitrary frustrations are more likely to increase aggressive drive than are nonarbitrary frustrations. One reason for this, he suggests, is that the arbitrary frustrations involve the nonfulfillment of an expectancy, and aggressive reactions are said to result primarily when the expected attainment of a goal is blocked. Bateson presented an essentially similar hypothesis in his discussion of cultural factors affecting the frustration-aggression relationship (1941). The Balinese apparently do not learn to expect strong satisfactions, supposedly because of the frequent frustrations imposed upon them by their mothers in the course of their socialization. Therefore, Bateson argued, "they are infinitely willing to suffer interruption" (p. 353).

Dollard and his colleagues generally did not consider the operation of expectancies in affecting the reactions to frustrations, and this lack is particularly felt in their analysis of frustrated response-sequences. There is no unequivocal support for their principle that the strength of the aggressive drive increases with the number of specific frustrations. The only evidence they give for the hypothesis is anecdotal, and experimental tests of the proposition have yielded inconsistent results. Otis and McCandless (1955) conducted an investigation with preschool children placed in an eight-trial frustration task. The *Ss* showed significant increases in "aggressive-dominant" behavior scores from the first four to the last four frustration trials, and showed reliable decreases in a class of behavior found to be incompatible with aggression, "submissive-complaisant" behavior. In contrast to this support for the proposition, McClelland and Apicella (1945) obtained negative findings in an experiment employing college students on only two frustration trials.

These opposing results can be reconciled by raising the possibility of differences in the development of understanding or expectations following the first frustrations. The person who anticipates further frustrations after the first disappointment may not be as upset when they occur as one who does not expect them. Conceivably, then, to the extent that the Otis-McCandless results are correctly interpreted as showing an increase in overt hostility with repeated frustrations, their *Ss* may not have expected these continued blockings of their goal attempts. The McClelland-Apicella *Ss*, being older and more sophisticated, presumably did develop these anticipations.

Findings consistent with this reasoning were reported in an earlier paper by the present writer (Berkowitz, 1960a). The *Ss* received two acts from their partners. The first of these was either friendly or unfriendly in tone and the second (in a factorial design), although of a qualitatively different nature, also was either friendly or hostile. The data showed that the partner's second act had its greatest effect on *S*'s attitude toward him when it was

[1] This research was supported by Research Grant M1540(C1) from the National Institute of Mental Health, United States Public Health Service. The invaluable assistance of Thomas Murphy, Woodrow Call, and Diane Diamond, who collected the present data, is gratefully acknowledged.

inconsistent with the earlier behavior. That is, if the second act was relatively aggressive in manner, it provoked the greatest increase in unfriendliness toward the partner (*P*) when it had been preceded by the friendly behavior. (However, this was but one case of a more general principle: the greatest change toward friendliness following the second friendly act occurred when this behavior came after hostility from *P*.) It may be that the *S*s were relatively indifferent to the aggressive (or friendly) second act when this behavior was consistent with the earlier behavior because such actions were more or less expected of *P*.

The first experiment in the present report employed a procedure virtually identical to that used in the earlier investigations, and essentially the same results should be obtained. Assuming that the initial act from *P* gave rise to an hypothesis about him, there should be a greater resulting change in attitude toward *P* following his second communication if this behavior was greatly inconsistent with the hypothesis than if the act was in accord with *S*'s conception of *P*.

This prediction appears to differ from the results obtained in a number of studies in the area of "person perception." Typically, investigators have found a "primacy effect" in which the first-presented information concerning a person presumably gave rise to a general hypothesis about him, with later information then being assimilated to this hypothesis (Asch, 1952; Kelley, 1950; Luchins, 1957). For example, Kelley led one group of *S*s to believe that a speaker was a "warm" person, while other *S*s were induced to develop the hypothesis that he was "cold." Later, when the speaker delivered his talk, both groups of *S*s evaluated the speech in a manner consistent with their particular hypotheses about him. The present prediction, instead of assuming a similar assimilation tendency, if anything, argues for a "contrast (or recency) effect." It essentially maintains that *P*'s second action toward *S* is contrasted with the initial expectation, and as a consequence, influences *S*'s final attitude toward *P* to a greater extent than *P*'s first behavior.

The major difference between the present studies and the Asch-Kelley-Luchins investigations involves the discrepancy between *S*'s initial hypothesis and the information subsequently provided about *P*. It seems fair to assume that this later information in the Asch-Kelley-Luchins studies was not regarded as very different from *S*'s first concept of *P*, either because the information was somewhat ambiguous in nature or because it truly was not greatly inconsistent with the hypothesis. In summarizing Luchins' findings, Hovland (1958) stressed the fact that the *S*s did not consider the second set of information about the stimulus person to be incompatible with the first presented material, and he suggested that this is a limiting condition for the Luchins results. The suggestion is entirely in keeping with the present argument. There is reason to believe that if the *S*s had perceived the incompatibility between the two sets of material, a contrast would have occurred in which the later material would have been judged as being even more different from the first presented information than was actually the case (for example, as defined by independent judges [Berkowitz, 1960b]). The probable result of this contrast is a recency effect; the later information has a greater impact in affecting the final attitude toward *P* than does the first information. This essentially was the finding in the investigator's previous experiment (1960a), and presumably it arose from just this process: a contrast effect, in which *P*'s second act is evaluated as being extremely different from his first communication (and/or different from the initial hypothesis formulated concerning *P*), giving rise to the dominant influence of this later behavior upon *S*'s attitude toward *P*.

EXPERIMENT I

Method

Procedure. With some exceptions probably not relevant to our present purposes, the procedure of Experiment I was identical to that employed in the earlier investigation (Berkowitz, 1960a). Summarizing this procedure, two *S*s of the same sex were brought together ostensibly for a study of the formation of "first impressions." After receiving the instructions and a brief description of their task, they went into separate cubicles where they indicated their impressions of their partners on the first questionnaire (Q1). Three intercorrelated items in this questionnaire index *S*'s baseline level of friendliness toward *P*. However, the questionnaire purported to assess *S*'s judgments of *P*'s personality.

The *S*s then wrote two notes to each other at 5-min. intervals in which they described themselves. They were to use their partner's self-descriptions in evaluating him on the second questionnaire. *E* substituted previously constructed notes for the notes actually written in order to produce the initial hypothesis concerning *P*. The notes received by the *S*s in the Initial Hostility condition indicated that the partner had little regard for the University and most of its students, while *S*s in the No Initial Hostility condition were led to believe that the partner liked life at the University, and was fairly active in student affairs. We assume that the former *S*s expected *P* to be an unpleasant or hostile person as a result of these notes, while the latter *S*s expected him to be friendly. Following this, *S*s again indicated their impressions of their partners on a second form of the questionnaire (Q2). The same items used in Q1 were employed in testing the success of the initial hypothesis-creating manipulation.

E then collected Q2 and told the *S*s that they would now be given still more information about their partners: they would be told what their partners thought of them. *E* gave each *S* a previously completed Q2, saying that this was the partner's second rating of the *S*. Half of the *S*s in each of the above conditions (Unfavorable Evaluation) received an evaluation with uniformly unfavorable ratings, with the others (Favorable Evaluation) getting highly favorable ratings. After the *S*s had completed the third questionnaire (Q3) and an adjective check-list, the deceptions and experimental procedures were explained and the *S*s were dismissed with a pledge of secrecy from them.

Statistical analyses and dependent variables. The *S*s in each condition were differentiated according to whether their Q1–Q2 changes indicated either a decrease or an increase in unfriendliness toward *P* following his initial communications. *S*'s initial hypothesis concerning *P* may be inferred from these changes. A decrease in unfriendliness is taken as denoting an Expectation of Friendliness, while an increase in this unfavorable attitude defines an Expectation of Hostility. Within each of these groups, the *S*s were further subdivided in terms of whether their inferred hypotheses were strong or weak. They are said to have a Strong Expectation if their attitude changes after *P*'s initial communications coincided in direction with the changes displayed by the majority of *S*s (and therefore, presumably were in accord with the objective nature of the notes). The *S*s exhibiting Q1–Q2 attitude changes differing in direction from the majority of *S*s' responses to the particular set of notes are defined as having a Weak Expectation.[2]

The unfriendliness score involved in these attitude changes is based upon the responses to three intercorrelated items on the 13-item questionnaire forms. The items ask, "Would you admit him (her) into your circle of close friends?"; "Would you like this individual as a roommate?"; and "Would you admit him (her) into the social organization of greatest importance to yourself?" In each case the answers consist of a checkmark placed anywhere on a line ranging from Definitely yes to Definitely no, with the score being the distance away from the Friendly end of the scale in one-quarter inch units summed over the three items.

In addition to these questionnaire responses, the *S*s at the end of the experimental session were asked to describe *P* on an adjective check-list. Sixty-seven of the adjectives on this list had been shown in previous research to be particularly sensitive to changes of feeling on the evaluation dimension. Using the classifications of 63 undergraduate judges as a basis for dividing the terms into Good or Bad traits, *S*'s final level of unfriendliness toward *P* was assessed by adding the number of bad traits *P* was said to have to the number of good characteristics he was said not to possess. These scores were then "normalized" by means of a square-root transformation. For both the rating scale and check-list measures, our working assumption is that negative evaluations (or unfriendliness responses) are indicative of hostility.

Subjects. There are 160 *S*s (half of them women) in this first experiment, recruited from introductory psychology courses at the University of Wisconsin. These *S*s were distributed evenly, maintaining a one-to-one sex ratio, among the four experimentally-created conditions involving the combination of notes and evaluations from *P*. A preliminary analysis of variance of the data employing these classifications did not reveal any significant sex differences, and this variable is therefore ignored in the present report.

Results

Effectiveness of the initial hostility manipulation. There were no significant differences among the eight conditions in the mean level of unfriendliness toward *P* on the first questionnaire ($F = 1.38$ with 7 df). The frequencies of *S*s in each of the conditions, shown in Table 1, demonstrate that the initial notes generally were successful in establishing the intended attitude toward *P*; by far the greatest proportion of *S*s changed their attitudes from Q1 to Q2 in a direction consistent with these communications. (When the various evaluation conditions are grouped, since the *S*s had not received these evaluations at this stage, $\chi^2 = 21.04$; $p < .001$).

The same conclusion can be drawn upon inspection of the condition differences in mean changes at this time. If only the *S*s who had increased in unfriendliness toward *P* are considered, there was a significantly greater increase in the *S*s receiving the hostile notes than in the *S*s getting the friendlier communi-

[2] This reasoning assumes that all *S*s recognized, to some extent at least, the "objective" nature of *P*'s first communication. (Evidence consistent with this assumption is provided below.) Therefore, those whose feelings toward *P* changed in a direction inconsistent with this communication are regarded as being less certain of their initial hypothesis about *P*; they presumably had detected some evidence contradicting their initial belief.

TABLE 1
MEAN CHANGES IN QUESTIONNAIRE UNFRIENDLINESS TOWARD *P*

	Initial Hostility		No Initial Hostility	
	Unfavorable Evaluation (A)	Favorable Evaluation (B)	Unfavorable Evaluation (C)	Favorable Evaluation (D)
I Changes from Q1 to Q2				
Ss who:				
(a) Increased in unfriendliness	(32)	(24)	(14)	(13)
from Q1 to Q2	3.98[a]	3.69	2.16	2.04
(b) Decreased in unfriendliness	(8)	(16)	(26)	(27)
from Q1 to Q2	−1.68	−1.91	−2.66	−2.45
II Changes from Q2 to Q3				
Inferred initial hypothesis:	(32)	(24)	(14)	(13)
	Strong-Hostility		Weak-Hostility	
(a)	0.92_{a}[b]	-3.09_{a}	2.00_{ad}	-0.84_{b}
Inferred initial hypothesis:	(8)	(16)	(26)	
	Weak-Friendly		Strong-Friendly	
(b)	2.95_{ab}	-1.24_{b}	3.26_{a}	0.90_{a}

[a] A positive score throughout signifies an increase in unfriendliness toward *P*.
[b] Cells containing the same subscript are not significantly different from each other (at the .05 level) by two-tailed *t* test.

cation (t_s for the Unfavorable Evaluation and Favorable Evaluation conditions are 4.14 and 3.34, respectively, both significant at $p < .01$).[3] The differences in attitude change were not as great in the case of the *Ss* who decreased in hostility toward *P* from Q1 to Q2, but again the *Ss* receiving the friendly notes tended to become somewhat friendlier (t_s = 1.75 and 1.23, respectively; the first difference is significant at the .10 level). These differences, it will be noted, are consistent with the assumption that attitude changes in keeping with the majority reaction to the messages denote a stronger expectation than changes deviating from the majority reaction. The degree of attitude change toward *P* is likely to be associated with the strength of the opinion *S* formed regarding *P*. Those who changed more may well have formed a more definite hypothesis.

Reactions to P's second acts. Generally, the Q2–Q3 attitude changes resulting from *P*'s evaluations of *S* follow the same pattern as the results of the preceding study. The smallest shifts in hostility toward *P*, whether toward increased friendliness or unfriendliness, occurred in those conditions where the nature of the evaluations received from *P* was consistent with the inferred strong expectations developed from his earlier act. Thus, definitely expecting hostility from *P*, the *Ss* in Condition A(a) showed a reliably *smaller* increase in unfriendliness after *P*'s unfavorable evaluation of them than the *Ss* in Condition C(b) who presumably strongly expected a friendlier act from him (t = 4.59; p = .001). They also exhibited a somewhat (but not significantly) smaller increase in unfriendliness toward *P* than the *Ss* in Condition C(a) who had also received the unfavorable evaluation from *P* but who had only tentatively expected hostility from him (t = 1.83; p = .10). This change in Condition A(a), however, was significantly less than that for the *Ss* in Condition A(b) who presumably had a weak hypothesis of friendliness from *P* (t = 2.48; p = .02).

Essentially the same pattern was displayed by the *Ss* receiving the friendlier favorable evaluations from *P*. This action probably was consistent with the definite hypothesis held by the *Ss* in Condition D(b), and they were much less likely to change toward friendliness following this favorable evaluation than the *Ss* strongly expecting hostility (i.e., those in Condition B(a); t = 4.18; p = .001). The B(b) *Ss*, only tentatively expecting friendli-

[3] The *t* tests (two tailed) conducted in Experiment I had their error terms based upon the "residual mean square" from a preliminary "among groups" analysis of variance. The tests of condition differences in Experiment II utilized the Duncan Multiple Range procedure (Duncan, 1955), which could be carried out because of the equal cell frequencies in this latter study.

TABLE 2
MEAN LEVEL OF UNFRIENDLINESS TOWARDS *P* ON THE ADJECTIVE CHECK LIST ADMINISTERED AT THE END OF EXPERIMENT I
(Square-Root Transformation)

	Initial Hostility		No Initial Hostility	
	Unfavorable Evaluation (A)	Favorable Evaluation (B)	Unfavorable Evaluation (C)	Favorable Evaluation (D)
Ss who:				
(a) Increased in unfriendliness from Q1 to Q2	(32) 5.86	(24) 4.77	(14) 4.65	(13) 3.72
(b) Decreased in unfriendliness from Q1 to Q2	(8) 5.44	(16) 3.61	(26) 4.59	(27) 3.63

ness from *P*, apparently were more pleased by *P*'s favorable evaluation of them than the former (D[b]) *Ss* who supposedly had a strong hypothesis of friendliness from *P* ($t = 3.63$; $p = .01$). Similarly, *Ss* in Condition D(a), tentatively expecting hostility from *P*, also became friendlier toward him upon receiving this pleasant act than did the D(b) *Ss* ($t = 2.80$; $p = .01$).

To summarize, these results clearly are in accord with the predictions developed on the basis of the earlier investigation. *Ss* receiving an act from *P* consistent with their presumed strongly held hypotheses about him showed relatively little change in attitude toward him. By and large, the greatest change in their feelings toward *P* was shown by *Ss* who received acts that deviated from these definite hypotheses.

Cumulative effects of frustration-sequences. The data presented up to this point have referred to attitude changes. In actuality, of course, the *Ss*' responses were in the form of absolute scores on the three-item rating scale. Inspection of the scores on the final questionnaire tends to confirm the *Frustration-Aggression* hypothesis regarding the cumulative effects of repeated frustrations. The *Ss* in Condition A, who had received two unfriendly acts from *P*, had the highest level of hostility towards him by the end of the experiment, although these scores were not reliably greater than those in Condition C.

Similar results were found with the adjective check-list administered at the end of the experimental session, as can be seen in Table 2. The *Ss* in Condition A(a), who had perceived two clear attacks upon them, described the attacker in a significantly more unfavorable manner than the *Ss* perceiving only one definite attack. (Comparing the former condition with B(a), $t = 4.94$, $p < .01$, and with Conditions C(a) and C(b), it is 3.78, $p < .01$, and 4.38, $p < .01$, respectively.)

Nevertheless, it is clear from the previous analysis that this final level of hostility in Condition A(a) might not have increased much further with additional frustrations. *S* conceivably could learn to expect the frustrations as they are repeated so that their effect is lessened. However, this is not to suggest that a long series of frustrations would necessarily produce ever-diminishing reactions with the development of these expectations. The person may reach a "breaking point" after continually failing to reach his goals, for example, as McClelland suggests (1951), when he fears he cannot at all cope with the situation confronting him. His reaction then could well be a great surge of anger toward the frustrator.

EXPERIMENT II

The next experiment was designed to test the possibility that contrast effects in judgments of behavior are less likely to occur when *S* has a very general hypothesis about *P*. Conceivably, the difference between the foregoing results and the Asch-Kelley-Luchins findings might stem from differences in hypothesis-width. The *Ss* in the latter investigations formed general impressions of the stimulus person, while *Ss*' initial expectations concerning *P* in Experiment I might have been very restricted in scope, pertaining more directly to *P*'s probable next act rather than to what he was like *in general*. If so, broadening this hypothesis to cover *P* as a person might result in an assimilation effect in which *P*'s second act is interpreted in a manner consistent with the hypothesis.

Method

Procedure. Male and female college students who had volunteered to serve in an experiment on "personality factors in the ability to assess others' personalities" first participated in a group testing session in which they

TABLE 3

MEAN CHANGES IN LEVEL OF UNFRIENDLINESS TOWARDS *P* ON THE ADJECTIVE CHECK LISTS

Experiment II

	Hostile Expectation		Friendly Expectation	
	Hostile Note	Friendly Note	Hostile Note	Friendly Note
First Impression: Absolute score on CL-1	5.6	9.0	8.3	6.2
Change after expectation created	9.3[a]	8.3	−2.7	−2.7
Change after note	-1.3_b[a]	-5.4_a	4.3_a	0.44_{ba}

Note.—$N = 9$ in each condition.

[a] In all cases a positive score signifies an increase in unfriendliness toward *P*, while a negative score denotes a decrease in this attitude.

[b] Cells containing the same subscript are not significantly different from each other (at the .05 level) by Duncan Multiple Range Test.

responded to a personality inventory. Some time later two *S*s of the same sex, but who did not know each other, were scheduled for a given laboratory hour. When they arrived they were reminded of the ostensible purpose of the study and then ushered into separate rooms. There, each *S* indicated his first opinions of the other person (*P*) on a 33-item adjective check-list (CL-1), supposedly in order to demonstrate the accuracy of his first impressions, but actually to provide a baseline level of friendliness toward *P*.

*S*s were then given instructions designed to create initial expectations regarding *P*. After saying he would give them more information about their partner, *E* showed each *S* the partner's "personality profile" supposedly prepared by professional psychologists on the basis of the earlier test results. This "profile" consisted of graphic ratings on five bipolar traits. On four of these traits (for example, "introversion," "self-insight," etc.), the ratings indicated *P* was in the "normal" or "average" range. The fifth rating was varied with the experimental condition. Half of the *S*s were led to believe *P* was "extremely friendly" (Friendliness Expectation), while the others were shown that he was "very hostile" (Hostility Expectation). Once *E* was sure that the ratings were understood he gave *S* an alternate form of the adjective check-list (CL-2), saying he was to evaluate the partner "on the basis of all the information he now possessed." The scores obtained at this time were used to test the success of the manipulation of expectation.

In the last part of the experiment *S* and his partner were to write a note to each other in which they were to talk about themselves. The *S*s were told the note from *P* would provide them with further information about *P*. In actuality, however, they were given the previously prepared communications employed to manipulate attitudes in the preceding study. Half of the *S*s in each of the above conditions received the message, now combined on one card, used to increase unfriendliness (Hostility Note), and the remaining *S*s received the friendliness-increasing communication (Friendly Note).

After the notes were written, "exchanged" and read, each *S* rated his partner again on an adjective check-list (CL-3) that was composed of both alternate forms employed before. Four of the nine *S*s in each experimental condition were males.

Dependent variables. As in the first experiment, the unfriendliness score obtained from the check-list was computed by summing the number of unfavorable traits attributed to *P* with the number of good characteristics he was said not to possess. In testing the effect of the expectancy-creating statements, the CL-1 score for each *S* was subtracted from his CL-2 score. The final score, denoting the change in unfriendliness created by the communication from *P*, was obtained by considering only those adjectives in the final check-list that had been used in CL-2. The CL-2 score was subtracted from this score to yield the measure of the reaction to *P*'s behavior.

Results

The statistical findings obtained with the adjective check-list measures of unfriendliness are summarized in Table 3. Analysis of variance of the change scores following the expectation manipulation indicate that this treatment was highly successful. *S*s given the hostile expectation increased considerably in the unfavorableness of their descriptions of *P*, while those given the friendly expectation tended to develop a somewhat more favorable image of him. (This difference is significant at $p < .01$).

The reactions to the communication from *P*, shown in the last row of the table, generally were consistent with the results of the preceding study. The hostile note created the greatest increase in unfriendliness when it was sent to the *S*s having the friendly expectation, while there was the greatest decrease in unfriendliness following the friendly note when the message was received by *S*s presumably not expecting this type of behavior.

The implication of these findings is clear. It is unlikely that the previous results can be attributed to narrow hypotheses restricted to *P*'s probable next act. The contrast effects consistently obtained in these studies apparently arise because the second (or the to-be-

evaluated) actions differ greatly from the actions expected by *S*.

DISCUSSION

The findings of these experiments can readily be understood as instances of judgmental phenomena. As is well known, when a person makes a judgment he essentially compares the stimulus with some standard, and a contrast effect occurs in this process to the extent that the stimulus differs from the standard (Berkowitz, 1960b). That is, *S* evaluates the difference as being greater than it really is (as defined, for example, by an observer not affected by the standard). In the present case, *S*'s expectation or hypothesis concerning *P* can serve as the standard against which *P*'s subsequent behavior is evaluated. When *P*'s action is altogether inconsistent with *S*'s hypothesis there apparently is an analogous contrast effect, and the action has a considerable impact upon *S*'s final attitude toward *P*. This judgmental analysis suggests, furthermore, that it may not be necessary to differentiate strong from weak hypotheses, a distinction, it will be recalled, introduced in Experiment I in connection with the correspondence between *S*'s and the majority reaction to *P*'s initial communication. For example, *S* was said to have only a tentative expectation concerning *P* if his attitude change toward *P* following this communication differed in direction from that exhibited by the majority of *S*s in his condition. Subsequent behavioral deviations from these presumably weak hypotheses did not have the same effect generally as deviations from supposedly stronger hypotheses. However, these perhaps unparsimonious interpretations concerning hypothesis strength can be avoided by making a somewhat different set of assumptions. Here we assume that the unfavorableness of *S*'s initial hypothesis about *P* can be coordinated directly to the changes in his questionnaire ratings of *P* from Q1 to Q2. Thus, the greater the change in an unfriendly direction, the more unfavorable is *S*'s hypothesis. If *P*'s second action be given an arbitrarily maximum score (say, 10) if it is aggressive in tone, and a score of 0 if it is friendly, the magnitude of the difference between this initial hypothesis and *P*'s subsequent behavior can then be computed. Extrapolating from judgmental research showing that the extent of the contrast effect is in direct ratio to the discrepancy between the stimulus and the standard, the greater the difference between the initial hypothesis and the later action, the greater should be *S*'s attitude change toward *P* following this action. Testing the prediction via the condition means shown in Table 1, the obtained correlation (tau) is +.71, which is significant at approximately the .03 level. The analogy therefore seems to hold. Unexpected behaviors may well be evaluated as being even more discrepant from the hypothesis than they really are, with the degree of the enhanced discrepancy a function of the degree of unexpectedness.

This reasoning has a direct bearing upon the hypotheses advanced by Dollard and his colleagues. Instead of supporting the Pastore-Bateson contention that it is necessary to distinguish among classes of frustrations because those that are unexpected are more likely than anticipated deprivations to produce aggressive reactions, the present findings can easily be incorporated into the Yale framework. According to Dollard, Doob, et al. (1939), the strength of the instigation to aggression is a positive function of the degree of frustration. We now can say that as a result of the contrast effect, unexpected frustrations are more likely to give rise to hostile responses because they probably are evaluated as being more severe than expected frustrations. In other words, there is a greater perceived interference with *S*'s goal attainment if the frustration is not expected than if it is anticipated because of the operation of judgmental processes upon the evaluation of the frustrating event.

The present analysis also questions the possibility raised by Miller and Campbell (1959) that there may be a "general primacy principle" which, at times, is masked by recency effects. The occurrence of primacy or recency may depend, in part at least, upon the perceived discrepancy between the first and later-presented information. The first material presumably gives rise to an hypothesis. In a sense, this expectation serves as the judgmental standard against which the later information is compared. If the later material appears consistent with the hypothesis (i.e., is close to the anchorage-expectation) it is incorporated into

the hypothesis, and a primacy effect is said to take place; the first information has had greater weight than the later information in influencing the final evaluation of the stimulus person, situation, or issue. On the other hand, if the later material is seen as deviating considerably from the hypothesis (and assuming it is not rejected), there is a contrast effect exaggerating the difference between the two sets of information. In this case, the later material may have a greater impact upon *S*'s final attitude toward the stimulus object, not only because it is fresher in his memory, as Miller and Campbell point out, but because of the contrast-enhanced evaluation of the material.

SUMMARY

Two experiments were devised to test predictions developed on the basis of an earlier investigation (Berkowitz, 1960a). In this study employing pairs of *S*s it was found that the partner's act had its greatest effect on *S*'s final attitude toward him when it was inconsistent with the partner's earlier behavior. For example, if the second act was relatively aggressive in manner it provoked the greatest increase in unfriendliness toward the partner (*P*) when it had been preceded by friendlier behavior. The first experiment reported here, employing a procedure virtually identical to that used in the previous study, obtained essentially similar results.

The second experiment, yielding the same pattern of findings, showed that these results cannot be attributed to "narrow" hypotheses on the part of *S*s in the first study, restricted to *P*'s probable next act.

These contrast effects appear to differ from other findings in the area of "person perception". The discrepancies are interpreted in terms of a judgmental analysis. This judgmental analysis was also employed to make the present findings consistent with the frustration-aggression hypotheses advanced by Dollard, Doob, et al. (1939). The present paper suggests that unexpected frustrations produce a stronger aggressive reaction than anticipated frustrations because, as a result of the hypothesized contrast effect, the former probably are evaluated as being more severe.

REFERENCES

ASCH, S. *Social psychology*. New York: Prentice-Hall, 1952.

BATESON, G. The frustration-aggression hypothesis and culture. *Psychol. Rev.*, 1941, **48**, 350–355.

BERKOWITZ, L. Some factors affecting the reduction of overt hostility. *J. abnorm. soc. Psychol.*, 1960, **60**, 14–21. (a)

BERKOWITZ, L. The judgmental process in personality functioning. *Psychol. Rev.*, 1960, **67**, 130–142. (b)

DOLLARD, J., DOOB, L., MILLER, N., MOWRER, O., & SEARS, R. *Frustration and aggression*. New Haven: Yale Univer. Press, 1939.

DUNCAN, D. Multiple range and multiple F tests. *Biometrics*, 1955, **11**, 1–45.

HOVLAND, C. I. The role of primacy and recency in persuasive communication, In Eleanor E. Maccoby, T. Newcomb, & E. Hartley (Eds.) *Readings in social psychology*. New York: Holt, 1958.

KELLEY, H. H. The warm-cold variable in first impressions of persons. *J. Pers.*, 1950, **18**, 431–439.

LUCHINS, A. S. Primacy-recency in impression formation. In C. I. Hovland (Ed.) *The order of presentation in persuasion*. New Haven, Conn.: Yale Univer. Press, 1957. Pp. 33–75.

MASLOW, A. H. Deprivation, threat, and frustration. *Psychol. Rev.*, 1941, **48**, 364–366.

MCCLELLAND, D. C., & APICELLA, F. A functional classification of verbal reactions to experimentally induced failure. *J. abnorm. soc. Psychol.*, 1945, **40**, 376–390.

MCCLELLAND, D. C. *Personality*. New York: Dryden, 1951.

MILLER, N. E. The frustration-aggression hypothesis. *Psychol. Rev.*, 1941, **48**, 337–342.

MILLER, N., & CAMPBELL, D. T. Recency and primacy in persuasion as a function of the timing of speeches and measurements. *J. abnorm. soc. Psychol.*, 1959, **59**, 1–9.

OTIS, N., & MCCANDLESS, B. Responses to repeated frustrations of young children differentiated according to need area. *J. abnorm. soc. Psychol.*, 1955, **50**, 349–353.

PASTORE, N. The role of arbitrariness in the frustration-aggression hypothesis. *J. abnorm. soc. Psychol.*, 1952, **47**, 728–731.

(Early publication received November 5, 1959)

Journal of Abnormal and Social Psychology
Vol. 60, No. 3, 1960

CRITIQUE AND NOTES

FRUSTRATION, AGGRESSION, AND MURDER[1]

STUART PALMER
University of New Hampshire

IN THE study reported here, 51 male murderers and their 51 nearest-age brothers were investigated with respect to frustration in infancy, childhood, and adolescence as an influencing variable (cf. Dollard, Doob, Miller, Mowrer, Sears, Ford, Hovland, & Sollenberger, 1939). Frustration was defined as lack of satisfaction of an innate or learned need. A murderer was defined as a person who has been legally convicted of first or of second degree murder or of first degree (non-negligent) manslaughter. The central hypothesis was: There is a significant positive relationship between the amount of frustration experienced by persons in infancy, childhood, and adolescence, on the one hand, and whether or not they later commit murder, on the other hand. Aggression release during the preadult years was considered a major control variable.

METHOD

The records of all males serving prison sentences for murder or first degree manslaughter in Massachusetts, New Hampshire, Maine, and Rhode Island during the fall of 1957 were investigated. In order to be used as cases in this study, it was necessary that, as well as having brothers, the murderers have mothers who were available for interviewing. Of 254 imprisoned men, 76 met the following Criteria: 1. under 51 years of age; 2. court records indicated that the subject was believed to have physically committed the murder; 3. at least one brother who had attained the age the murderer was when he committed murder; 4. mother living; 5. mother able to answer questions in English; and 6. mother residing within 150 miles of Durham, New Hampshire. Twenty-five of the 76 mothers were not available for interviewing, leaving 51 *Ss* in the murderer group. Where a murderer had more than one brother who fitted Criterion 3 above, the brother nearest the murderer in age was chosen, to yield a control group of the same size. None of these control brothers was known to have committed murder. Use of the brothers as control *Ss* made it possible to hold relatively constant such factors as age, sex, social class of parents, religion, and the like.

[1] Supported in part by four grants from the Central University Research Fund of the University of New Hampshire.

Some data, such as intelligence quotients, were obtained from the prison records of the murderers. Most were obtained through interviewing the mothers with the use of a questionnaire composed of 143 items, a majority of which were close-ended. Each mother completed the questionnaire twice, once for the murderer son, once for the non-murderer. The order of questioning about sons was reversed with alternate mothers.

Indices of Physical Frustration, Psychological Frustration, General Frustration, Socially Acceptable Aggression Release, Socially Unacceptable Aggression Release, and General Aggression Release were developed. In the Index of Physical Frustration, the *S* was allotted one point for the following:

1. A difficult birth
2. Serious effects of forceps at birth
3. Each serious operation
4. Each serious illness
5. Each serious accident
6. Each serious beating by someone other than a parent or step-parent
7. One or more serious beatings by the mother; and
8. One or more serious beatings by the father or step-father; with respect to this index as well as to those below, words such as difficult and serious were operationally defined

In the Index of Psychological Frustration, one point was allotted the *S* for the following:

1. Each serious, physical defect that could give rise to social disapproval
2. Very rigid time schedule adhered to by the mothers in caring for the child during his first two years
3. Bottle-feeding during some portion of the first two months of life
4. Fixed schedule followed by mother in feeding the child during the first two months
5. Final weaning started before age 12 months
6. Toilet-training begun before age 12 months
7. Sexual self-play stopped by mother before age 12 months
8. Mother responded to child very often by crying during first five years of his life
9. Mother responded to child very often by isolating herself during first five years of his life
10. Each trauma due to natural events or other individuals, excluding beatings
11. Great difficulty learning to read or never learned

12. Persistent bed wetting
13. Persistent stuttering; and
14. Spent a majority of time playing alone at age five

An *S*'s score on the Index of General Frustration was found by adding together his scores on the Index of Physical Frustration and the Index of Psychological Frustration.

In the Index of Socially Acceptable Aggression Release, one point was allotted the *S* for the following:

1. Verbal loudness at age five classified as loud or average
2. Anger short of temper tantrums at age five once a week or more
3. Reasonable verbal release at age five which occurred half or more of the times when angry
4. Physical fighting at age five once a week or more
5. "Talking back" to the mother at age five occasionally or more
6. Dirtying clothes at age five, half or more of the times when "dressed up"
7. Participation in athletics at age 17 occasionally or more
8. Hunting at age 17 occasionally or more; and
9. Mildly antagonizing others by verbal means at age 17 occasionally or more

In the Index of Socially Unacceptable Aggression Release, one point was allotted the *S* for the following:

1. Lying to the mother at age five occasionally or more
2. Stealing at age five occasionally or more
3. Swearing at age five occasionally or more
4. Temper tantrums once or more during childhood
5. Deliberately and seriously hurting animals once or more during childhood or adolescence
6. Fist-fighting once or more at age 17
7. Violent aggression other than fist-fighting when intoxicated, before age 20

Whether a behavioral form indicated socially acceptable or unacceptable aggression release was determined arbitrarily by the investigator. An *S*'s score on the Index of General Aggression Release was found by summing his scores on the Index of Socially Acceptable Aggression Release and the Index of Socially Unacceptable Aggression Release.

RESULTS

As Table 1 shows, significantly more murderers than control brothers scored high on the Indices of Physical, Psychological, and General Frustration. Further, when General Frustration scores within pairs of murderers and brothers were compared, it was found that in 42 pairs (82%) the murderer had

TABLE 1

SCORES OF MURDERERS AND CONTROL BROTHERS ON INDICES OF FRUSTRATION AND AGGRESSION

Score	Physical Frustration (No. of)		Psychological Frustration (No. of)		General Frustration (No. of)		Acceptable Aggression (No. of)		Unacceptable Aggression (No. of)		General Aggression (No. of)	
	M	B	M	B	M	B	M	B	M	B	M	B
0	4	13	2	4	0	2	5	2	22	35	5	1
1	5	13	2	11	1	4	10	5	12	12	6	3
2	8	12	6	13	3	10	9	12	6	3	7	15
3	5	7	10	11	2	7	16	11	8	1	7	7
4	3	5	7	4	4	5	7	4	2		9	5
5	6	0	5	5	7	8	3	4	1		7	4
6	3	1	6	2	2	6	1	8			2	10
7	7		6	1	2	3		5			4	2
8	6		2		5	3					2	3
9	3		1		5	2					2	0
10	0		4		1	1						1
11	1				1							
12–21					18							
$\bar{X}$	4.5	1.7	4.7	2.5	9.2	4.2	2.4	3.5	1.2	0.4	3.7	4.0
Chi-square	33.10		22.10		30.36		13.13		12.30		13.76	
P	<.01		<.01		<.01		<.05		<.01		<.05	
Scores combined for chi-square	5–11		0–1; 7–10		0–2; 3–4; 5–6; 7–8; 9–10; 11–21		0–1; 5–7		3–5		0–1; 7–10	

Note.—M designates murders; B designates control brothers. In each column, $N = 51$.

a higher score than his brother. In five pairs the scores were equal, and in four, the brother's score was higher.

The data in Table 1 also show that a significantly greater number of control brothers than of murderers had high scores on the Index of Acceptable Aggression Release. Conversely, a significantly greater number of murderers than of brothers had high scores on the Index of Unacceptable Aggression Release. On the Index of General Aggression Release, scores were approximately the same for the murderers as for the brothers (in spite of the fact that the Chi square value was significant at the .05 level).

Although detailed data are not included here, the relationship between general frustration and murder was investigated while holding general aggression release approximately constant. This was done by dichotomizing the total group of 102 subjects into two equal sub-groups; one, composed of 51 murderers and brothers who used general aggression releases extensively (score of 4 or more on the Index of General Aggression Release), and the other composed of 51 murderers and brothers who used general aggression releases infrequently (score of three or less). Frustration was positively associated with murder whether aggression releases were used extensively or infrequently. However, the association was statistically significant only when aggression releases were employed extensively.

DISCUSSION

Clearly, there are three especially major drawbacks in this study. First, the mothers may not have reported the facts correctly. Second, the group of murderers is probably not representative of any larger murderer population. Third, the indices of frustration and aggression release are crude; they do not take into account the fact that one incident of frustration or aggression release may not equal the weight of another such incident. However, as an exploratory study where more satisfactory data are difficult to obtain, it indicates the possibility of a functional relationship between preadult frustration and later presence or absence of murder.

REFERENCE

DOLLARD, J., DOOB, L. W., MILLER, N. E., MOWRER, O. H., SEARS, R. R., FORD, C. S., HOVLAND, C. I., & SOLLENBERGER, R. T., *Frustration and aggression*. New Haven: Yale Univer., 1939.

(Received June 11, 1959)

Journal of Abnormal and Social Psychology
Vol. 60, No. 3, 1960

SOME PHYSIOLOGICAL CORRELATES OF RORSCHACH BODY IMAGE PRODUCTIONS[1]

A. D. DAVIS[2]
University of Texas

RECENT investigations have given rise to the contention that some aspect of personality organization reflected in Rorschach percepts is involved in the selection of psychosomatic symptomatology (Cleveland & Fisher, 1954; Fisher & Cleveland, 1955, 1956, 1957). This personality variable is concerned with the individual's perception of his body exterior as having (or lacking) protective characteristics. The relative strength of this "body-image perception," according to Fisher and Cleveland (1957), can be quantified as a "barrier score" by an appropriate scoring procedure. This barrier score has been related to the projective performance of clinical groups manifesting specific psychosomatic symptom complexes, and it has been argued that these bodily perceptions predispose normal individuals to symptoms at either an "interior" or "exterior" bodily site through heightened physiological response tendencies (Fisher & Cleveland, 1955, 1957).

This formulation suggests that the greater

[1] This article is based on a dissertation submitted to the faculty of the department of psychology of the University of Texas in partial fulfillment of the requirements for the PhD degree. The author is indebted to Philip Worchel, Seymour Fisher, and Sidney E. Cleveland for their support, aid, and supervision in this study.

[2] Now at the Department of Psychology, Western State Hospital, Ft. Steilacoom, Washington.

number of Rorschach responses which emphasize "barrier" (impermeable, enclosing, or otherwise protective) surface qualities, the greater is the physiological reactivity toward the periphery of the body (skin and musculature). Conversely, the relative absence of these productions is related to increased physiological reactivity at an interior body site. The present study was formulated as a means of investigating this hypothesized relationship in normal subjects.

METHOD

Subjects

Group Rorschachs were administered to three sophomore psychology classes totaling 280 students at the University of Houston. Of these *Ss*, only the 146 males were utilized because of the nature of the physiological recording. From *Ss* having more than the median number of barrier responses, 25 were randomly selected and designated as the "high barrier" group (Group I); of *Ss* with less than the median number of barrier responses, 25 were randomly selected and designated as the "low barrier" group (Group II).

Apparatus and Physiological Measures

Physiological indices were obtained by the following methods and apparatus:

1. *Electromyogram* (EMG). Recordings of muscle action potential were obtained from two EMG electrodes fixed by a head band placed two inches apart over the frontalis muscles of the left eye. The muscle potentials were fed into an electronic current integrator after the design of Jacobson (1951) and this integral recorded on a standard Grass III-C, 7 channel, ink-writing oscillograph.

2. *Skin Resistance* (GSR). A measure of skin resistance was obtained from two 2-inch sponge electrodes placed on the dorsal and palmar surface of the left hand. The basic circuit of the GSR unit utilized for recording is reported by Davis and Porter (1931).

3. *Blood Pressure.* Systolic (SBP) and diastolic (DBP) blood pressures were obtained by the cuff and auscultatory method from the right arm. Mean blood pressure (MBP) was computed as the average of the systolic and diastolic pressures.

4. *Ballistocardiograph Variables* (BCG). Four ballistocardiograph indices were utilized after the fashion of Funkenstein, King, & Drolette (1954): (*a*) An absolute count of pulse rate (P). (*b*) The mean height of the BCG-IJ wave was determined over the arbitrary time span of the IJ waves utilized to define stroke volume (IJ). The IJ wave-form characteristic in the BCG record is the most conspicuous characteristic of the record and it is generally agreed to be the headward ballistic reaction of the body to ventricular systole (Dock, Mendelbaum, & Mendelbaum, 1953). This variable has been empirically related to ventricular stroke volume by Starr, Schnabel, Mayock, & Austin, 1952). (*c*) The average height of the IJ unit was multiplied by the pulse rate over unit time (IJ × P). This measure was empirically related to cardiac output by Nickerson, Warren, & Brannon, 1947). As outlined by Funkenstein (1954), these variables were utilized without the empirically derived constants of Starr and Nickerson (1952) as purely relative physiological indices with no interpretation of absolute physiological meaning or value intended. (*d*) The mean blood pressure divided by the average height of the IJ waves multiplied by pulse defined peripheral resistance (MBP/IJ × P). This measure has similarly been empirically related to the total peripheral resistance of the cardiovascular system.

The instrument used to obtain these records was a critically damped accelerometer consisting of a two-gram weight pendulum which was suspended from a magnetic phonograph cartridge mounted in a block of wood 18 × 4 × 2 inches. The pendulum weight was immersed in a glycerin bath which prevented reverberation of the pendulum, thus giving clean records of the body ballistics. The accelerometer was designed to give identical wave form characteristics to that of the Virgil Hill Electromagnetic BCG (Dock, Mendelbaum, & Mendelbaum, 1953). The apparatus was placed across the legs at the ankles and recorded on the Grass oscillograph.

From these measures, two groups of three variables each were assigned to exterior and interior classifications. The rationale for this division was dependent upon the bodily site at which the greatest variation of the measures occurred. The exterior variables and the physiological process to which they are related were as follows: (*a*) muscle potential (EMG), (*b*) skin resistance (GSR), and (*c*) peripheral resistance (MBP/IJ×P).

The interior measures included (*a*) pulse rate (P), (*b*) stroke volume (IJ), and (*c*) cardiac output (IJ×P). Systolic and diastolic blood pressures (SBP) and (DBP) were obtained in order to compute the mean blood pressure (MBP).

According to the formulation, Group I (high barrier) individuals should show greater responsivity in the exterior group of variables, and Group II should reflect relatively greater physiological response on the measures designated as interior.

Procedure

Each *S* was instructed to lie down on a one-half inch horsehair pad on the floor. A sandbag was placed beneath his heels to lock his knees for the ballistocardiograph recording, and the remainder of the equipment was affixed while the investigator

casually engaged him in conversation about neutral topics. *S* was then instructed to relax and rest while the investigator obtained "recordings of his heart." An initial rest period of 20 min. was given in an attempt to establish conditions which were as basal as possible for resting-level measures. During the last 10 min. of the rest period, GSR readings were recorded, and the blood pressure was sampled three times. At the end of the rest period, a 2-min. sample recording of EMG and BCG was obtained and utilized as rest level measures. At this time, another sample of GSR and blood pressure was taken and averaged with the initial measures as an estimate of resting level.

Six arithmetic problems from the Wechsler Intelligence scale were read from cards. *S* was given only 30 sec. for each solution. The first two answers were accepted as correct and the last four as incorrect, regardless of their nature. The problems were read readily, and the investigator's standard comments after each "failure" were designed to be increasingly more depreciating. The GSR reading was noted after presentation of each problem, and the maximum fall in resistance during the problem period was considered as the maximum stress response on that variable. After presentation of the last problem, blood pressure was immediately taken, and a sample recording of BCG and EMG was obtained. These measures were utilized as the poststress (or response to stress) measures.

RESULTS

The scores of the combined groups on each variable were linearly transformed into distributions having a mean of 50 and a standard deviation of 10, making it possible to compare the individual's performance on any variable with his performance on any other relative to the sample mean of the variable. As a means of evaluating the differences in the groups' resting levels, *t* and *F* tests of significance were performed between groups on each of the variables in the pre-stress condition. The significant data are presented in Table 1. Marked differences occurred between the two groups on their resting measures. Group I was significantly higher on one exterior measure (EMG) at the .02 level with a one-tailed test. The Group II resting response level, as predicted, tended to be higher than that of Group I on the interior measure of P. However, the outcomes for IJ and IJ × P were reversed with Group I having the higher mean. Interestingly, the observed blood pressure measure, SBP and MBP, approached significance in favor of Group II and appeared to have greater variance than any other Group II performance. Similarly, Group I variance on the blood pressure measures was among the smallest of that group's performances. Inasmuch as the standard deviation for the distribution of the groups combined is equal to 10, discrepancies from this value are some relative indication of the lability of the physiological variable for the group.

TABLE 1

TRANSFORMED PRESTRESS MEANS, STANDARD DEVIATIONS, AND SIGNIFICANCE OF DIFFERENCES

Physiological Index	Group I		Group II		Mean Diff.	*t*	*P* (1-tailed)	*F*
	Mean	*SD*	Mean	*SD*				
Exterior								
EMG	53.06	9.71	46.94	9.50	6.12	2.26	.02	—
GSR	51.42	7.57	48.37	9.64	—	—	—	—
MBP/IJ × P	49.22	11.10	51.00	9.10	—	—	—	—
Interior								
P	47.97	10.78	52.03	8.90	4.06	1.45	.08	—
IJ	52.56	11.06	47.18	9.15	5.38	1.87	.045	—
IJ × P	51.11	11.73	48.41	7.18	—	—	—	2.66
Blood Pressure								
SBP	47.95	8.51	52.45	10.37	4.50	1.60	.09	—
DBP	48.92	8.85	51.08	11.07	—	—	—	—
MBP	47.50	8.46	52.49	10.93	4.99	1.81	.05	—

To make comparisons between the groups on their relative physiological reactivity to stress, pre- and postdifference scores were corrected for differences in initial resting level for a regression analysis. The procedure was recommended by Lacey (1956), who pointed out that the high correlation between resting and stress levels predisposes percentage or absolute-change data toward reflecting initial differences. The result of this treatment yielded a distribution of "Autonomic Lability Scores" free of dependence on base level and expressed in units which make it possible to compare relative responsivity on a number of different physiological variables.

Table 2 summarizes the data for the Autonomic Lability Scores. It is apparent that Group I responded to the stress situation in terms of the exterior variables with greater EMG, GSR, and MBP/IJ×P values than those of Group II. Although the GSR difference was not significant, the difference was in the predicted direction. With a one-tailed test, EMG and MBP/IJ×P may be considered as significant at the .01 and .06 levels

TABLE 2

AUTONOMIC LABILITY SCORE MEANS, STANDARD DEVIATIONS, AND SIGNIFICANCE OF DIFFERENCES BETWEEN GROUPS

Physiological Index	Group I		Group II		Mean Diff.	t	P (1-tailed)	F
	Mean	*SD*	Mean	*SD*				
Exterior								
EMG	53.39	9.34	46.81	9.60	6.58	2.46	.01	—
GSR	51.15	7.76	49.49	5.74	1.66	.86	.20	—
MBP/IJ × P	51.66	8.58	47.49	9.65	4.17	1.62	.06	—
Interior								
P	47.37	8.66	52.62	9.72	5.25	2.03	.02	—
IJ	47.87	7.56	53.08	11.05	5.21	2.17	.03[a]	2.13
IJ × P	47.81	6.74	52.25	8.30	4.44	2.07	.02	—
Blood Pressure								
SBP	45.89	9.20	53.73	9.26	7.64	3.00	.003	—
DBP	48.66	10.84	51.27	8.54	2.61	.94	.17	—
MBP	46.30	10.15	53.98	7.72	7.68	3.01	.005	—

[a] Two-tailed test due to significant F ratio.

TABLE 3

RANK ORDER OF RESPONSIVITY AND COEFFICIENTS OF THE TWO GROUPS ON EACH OF THE EXTERIOR AND INTERIOR VARIABLE IN PRE- AND POSTSTRESS AND AUTONOMIC LABILITY MEASURES

Physiological Index	Prestress Ranks		Poststress Ranks		Prepost Autonomic Lability Scores Ranks	
	Group I	Group II	Group I	Group II	Group I	Group II
Exterior						
EMG	1	6	1	6	1	6
GSR	3	4	2	5	3	4
MBP/IJ × P	5	2	4	4	2	5
Interior						
P	6	1	6	1	6	2
IJ	2	5	3	3	4	1
IJ × P	4	3	5	2	5	3
rho:[a]	−1.00		−.86		−.83	
Standard error	—		.111		.133	

[a] rho of .829 needed at .05 level of significance; .943 needed at .01 level.

respectively. Conversely, Group II responded in terms of the interior variables with P (t of 2.03), IJ (t of 2.17), and IJ×P (2.07) all significant beyond the .02 level. A one-tailed test was not reported on the IJ variable because of the significant difference in variances. The blood pressure measures, SBP and MBP, as in the resting level, were greater (.01) in Group II.

With the raw data transformed in such a way as to permit the relative order of responsivity across variables to be determined, the mean rank of each group on each variable was found and rank order (rho) correlations computed. These ranks and their rho coefficients for the prestress, poststress, and pre- and poststress regression analyses (Autonomic Lability Scores) are presented in Table 3. As the table suggests, there was a marked contrast in the form of a high inverse relationship. There was a perfect negative correlation between the order of Group I and Group II in the resting level; for the poststress and difference analyses, the rhos of −.86 and −.83 respectively were both significant beyond the .05 level.

DISCUSSION

The results indicate that the differences at resting level between the groups are not completely in accord with the hypotheses. For Group I (high barrier), however, the EMG measure is greater than that of the Group II (low barrier) measures. Similarly, for Group II, P is greater than in Group I with no evidence of differences between the groups on GSR, MBP/IJ×P, or IJ×P. It will be noticed that IJ×P (cardiac output) enters into both of the nonsignificant cardiac variables. It appears from inspection of the underlying variables (P and IJ) that the cardiac output, although not different, may be achieved in different ways during the resting measures. In Group I, it seems to be primarily a function of significantly higher stroke volume (IJ) and slow pulse; whereas in Group II, the cardiac output appears to result from low stroke volume and a significantly more rapid pulse. Additionally, the pre- to poststress means suggest that all cardiac activity in Group I decreases from rest to stress. This is undoubtedly related to the increase in this group's peripheral resistance, which would tend to impede any cardiac output. Nevertheless, cardiac activity in Group II seems excessive under stress in comparison to the increase shown by Group I *Ss*. On the other hand, the cardiac organization of Group I appears to becomes more efficient under stress, possibly through a slowing of the pulse, allowing greater ventricular filling before systole. These observations reflect highly significant differences in the organization of the two groups' cardiovascular functioning. Such characteristic differences in cardiovascular organization across individuals

were similarly reported by Lacey (1950, 1953) with *Ss* under stress.

To summarize the resulting physiological hierarchies of response of the two groups under stress: Group I (high barrier) reacts in terms of increased muscle potential and peripheral vasoconstriction, a decreased skin resistance, increased stroke volume of the heart, increased cardiac output, and increased pulse rate in that order. The Group II (low barrier) hierarchy of response under stress consists of an increase of stroke-volume, pulse rate, and cardiac output, and a decrease in GSR, increased peripheral resistance, and muscle potential in that order. These hierarchies clearly demonstrate a primacy of response in the exterior indicies for the high barrier group and for the interior indices in the low barrier group when both groups are under stress. These data tend to lend support for the hypotheses of this study if the hypotheses are restricted to considerations of the stress data. The resting level data do not support the contention that these physiological predispositions are characteristic of the two groups under basal conditions.

SUMMARY

The relationship between surface characteristics of Rorschach responses and physiological measures was investigated in order to test the hypothesis that the projected perception of one's body is related to the patterning of physiological responses. Nine physiological measures were obtained under rest and experimentally induced stress from 50 *Ss* selected at random from a larger sample of 146 normal male subjects. Physiological indices related to bodily exterior (muscle potential, skin resistance, cardiovascular peripheral resistance) and bodily interior (pulse rate, stroke volume, cardiac output) physiological responsivity were abstracted from the nine measures. The results indicate that the tendency to give impermeable or protective characteristics to Rorschach percepts is related to heightened physiological reactivity at an exterior body site and that the absence of these characteristics in Rorschach productions is related to heightened physiological response on interior indices of physiological response under stress.

REFERENCES

CLEVELAND, S. E., & FISHER, S. Behavior and unconscious fantasies of patients with rheumatoid arthritis. *Psychosom. Med.*, 1954, **16**, 327–334.

DAVIS, R. C., & PORTER, J. M., JR. A measuring device for the galvanic skin reflex. *J. gen. Psychol.*, 1931, **5**, 115–120.

DOCK, W., MENDELBAUM, H., & MENDELBAUM, R. A. *Ballistocardiography*. St. Louis: C. V. Mosby, 1953.

FISHER, S., & CLEVELAND, S. E. The role of body image in psychosomatic symptom choice. *Psychol. Monogr.*, 1955, **69**, (17, Whole No. 402).

FISHER, S., & CLEVELAND, S. E. Relationship of body image to site of cancer. *Psychosom. Med.*, 1956, **18**, 304–310.

FISHER, S., & CLEVELAND, S. E. An approach to physiological reactivity in terms of a body-image schema. *Psychol. Rev.*, 1957, **64**, 26–38.

FUNKENSTEIN, D. H., KING, S. H., & DROLETTE, M. The direction of anger during a laboratory stress-inducing situation. *Psychosom. Med.*, 1954, **16**, 404–413.

JACOBSON, E. Muscular tension and the estimation of effort. *Amer. J. Psychol.*, 1951, **64**, 112–117.

LACEY, J. L. Individual differences in somatic response patterns. *J. comp. physiol. Psychol.*, 1950, **43**, 338–350.

LACEY, J. The evaluation of autonomic responses. Toward a general solution. *Ann. NY Acad. Sci.*, 1956, **67**, 123–164.

LACEY, J. I., BATEMAN, DOROTHY E., & VAN LEHN, RUTH. Autonomic response specificity. *Psychosom. Med.*, 1953, **15**, 8–21.

NICKERSON, J. L., WARREN, J. V., & BRANNON, E. S. Cardiac output in man: Studies with the low frequency, critically-damped ballistocardiograph and method of right atrial catheterization. *J. clin. Invest.*, 1947, **26**, 1–10.

STARR, I., SCHNABEL, T. G., JR., MAYOCK, R. L., & AUSTIN, J. H. Relation of aortic pulse pressure curves both to left ventricular stroke volume and to output at every instant of systole. *Fed. Proc.*, 1952, **11**, 152.

(Received October 7, 1958)

Journal of Abnormal and Social Psychology
Vol. 60, No. 3, 1960

MANIFEST ANXIETY AND PRESTIMULUS CONDUCTANCE LEVELS

GERARD J. McDONNELL AND JOHN A. CARPENTER
Yale University

NUMEROUS studies reporting correlations between conductance or palmar sweating changes and changes in anxiety stimuli or self-report of tension have encouraged the view that conductance level, from which the changes have been noted, represents the anxiety or tension[1] level of an individual. But studies attempting to determine the function relating conductance levels and manifest anxiety are rare. Hence, it is a dubious procedure to infer that one individual is more anxious than another on the evidence that his conductance levels are higher. Moreover, lacking knowledge of the function between anxiety and conductance level, it is, strictly speaking, improper to interpret changes in level as unequivocal evidence of changes in anxiety.

The following study was designed to clarify this problem. Specifically it was planned to obtain data on two questions:

1. If Individual A has the same prestimulus level as B, do the two have the same manifest anxiety level?

2. If A has a different level than B, does the one with the higher conductance also have the higher anxiety level?

METHOD

Thirty-two men, all graduate students, and eight women, students or students' wives, aged 21 to 28, served as paid subjects (*Ss*).

While *Ss* played cards, electrodes were fastened to the plantar surfaces of both feet. Essentially simultaneous recordings were taken from groups of eight *Ss* by a method described elsewhere (McDonnell & Carpenter, 1959). Recordings lasted for 15 min. for each of the two sessions (spaced one week apart). The individual's mean conductance level for both sessions was taken as his prestimulus conductance level. Manifest anxiety level was obtained by administration of the Mandler-Sarason General Anxiety scale at the second session.

RESULTS AND DISCUSSION

The questions prompting this inquiry were (*a*) if A and B have the same conductance level, do they report the same anxiety, and (*b*) if B has a higher level than A, does he also report greater anxiety?

Answers are obtained by referring to the relation between anxiety and conductance shown in Fig. 1. Each plotted point represents an average of the conductance level of those *Ss* whose anxiety scores were found in a given anxiety interval. The correlation ratio was .722 ($P < .01$ with 39 *df*).

Thus, it appears that if A and B both have high conductance levels, they are apt to report moderate anxiety; but if they have low conductance levels, *S* could report either high or low anxiety. A similar difficulty obtains in predicting differences in anxiety from the fact that B has a higher conductance than A. The higher conductance signifies higher anxiety only when both *Ss* report relatively low anxiety. If both report high anxiety, the higher conductance would be expected in the *S* with lower anxiety.

The evidence that low conductance may be associated with high anxiety corresponds to two previous observations. Van der Valk and Groen (1950) found that *Ss* with diseases supposedly symptomatic of high tension (ulcerative colitis, peptic ulcer, diabetes, etc.) showed lower conductance levels than a group of nondiseased normals. Similarly, Mowrer, Light, Luria, & Seleny (1953) reported an overtly high anxious patient who had "phenomenally" low levels of palmar sweating. The idea that similar conductance levels can characterize *Ss* with different anxiety levels is supported by the same author, who noted that individuals differing widely in tension level showed nearly identical sweating levels.

Greater awareness of anxiety may be accompanied at first by greater autonomic activity. Thereafter, some process, possibly an extension of homeostatic restraint as conceived by Lacey (1956), intervenes to modulate sympathetic effects. The result is a lowering of conductance levels.

Although this study only furnishes evidence on the meaning of conductance *levels*, the finding of a curvilinear function suggests caution in interpreting conductance *changes*. Depending on *S*'s level of anxiety, a decrease in conductance could signify an increase in anxiety for one *S* and just the reverse for another. Illustratively, Mowrer found with some patients that a decline in sweating was accompanied by an increase in reported tension. Similarly, as other investigators (Van der Valk & Groen, 1950) have interpreted their results, an increase in conductance could, in some *Ss*, mark a decrease in tension.

[1] The study assumes, without proof, an equivalence between "anxiety" and "tension." "Anxiety" will be used hereafter for both.

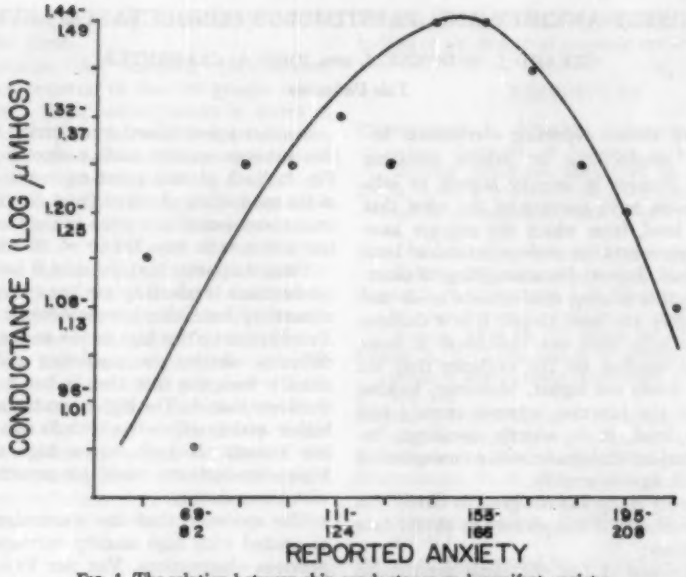

FIG. 1. The relation between skin conductance and manifest anxiety.

The results of this study require two qualifications: size and nature of sample, and the validity of the anxiety scale.

Sample size is important where correlation ratios are concerned, because small frequencies in the grouping intervals exaggerate the coefficient. Although care was taken to use gross categories, the number of *S*s at the extremes of the distribution were nevertheless small. Again, the twin circumstances of being voluntary and paid might have selectively influenced both conductance and anxiety score. However, the absence of satisfactory norms for conductance or anxiety levels makes identification of such influences impossible. Including results of both sexes seemed warranted for two reasons: females resembled the males in range of both anxiety scores and conductance levels, and they exhibited the same curvilinear relation between the two variables.

While reliability of the Anxiety scale is satisfactory, validity is not so well established (Mandler & Sarason, 1952). The lower anxious group was found here to be more heterogeneous in conductance level than the higher anxious group, possibly indicating that such low scores include both genuinely low anxious *S*s and *S*s who are denying or suppressing anxiety. It is equally possible that the greater variability of this group results from (*a*) weakness of discrimination in the Anxiety scale or (*b*) lack of correspondence between anxiety and conductance at the low end of the anxiety continuum. The data do not permit a decision between these alternatives.

SUMMARY

This study investigated the functional relationship between conductance levels and an index of manifest anxiety. A curvilinear relationship was found and discussed in relation to the use of conductance level as an index of emotionality.

REFERENCES

LACEY, J. I. The evaluation of autonomic responses: Toward a general solution. *Ann. NY Acad. Sci.* 1956, **67**, 123–164.

MCDONNELL, G. J., & CARPENTER, J. A. Anxiety, skin conductance and alcohol. *Quart. J. Stud. Alcohol*, 1959, **20**, 38–52.

MANDLER, G., & SARASON, S. B. A study of anxiety and learning. *J. abnorm. soc. Psychol.*, 1952, **47**, 166–173.

MOWRER, O. H., LIGHT, B. H., LURIA, Z., & SELENY, M. P. Tension changes during psychotherapy with special reference to resistance. In O. H. Mowrer (Ed.), *Psychotherapy*. New York: Ronald, 1953.

VAN DER VALK, J. M., & GROEN, J. Electrical resistance of the skin during induced emotional stress. *Psychosom. Med.*, 1950, **12**, 303–314.

(Received July 23, 1958)

Journal of Abnormal and Social Psychology
Vol. 60, No. 3, 1960

A PSYCHOLOGICAL FACTOR APPARENTLY ASSOCIATED WITH NEOPLASTIC DISEASE[1]

LAWRENCE LESHAN
The Institute of Applied Biology, New York
AND MARVIN REZNIKOFF
The Institute of Living, Hartford, Connecticut

AN INVESTIGATION of 250 patients with malignant disease (LeShan, 1956) indicated a pattern in 62% of this group and in 10% of the controls which included a psychologically traumatic event early in life and a marked amount of disturbance in the individual's ability to relate to others. A study by Reznikoff (1955) of women with cancer of the breast similarly revealed that a significantly higher proportion of cancerous women than controls had had a sibling die in childhood.

It was decided to test the hypothesis that the cancer patient has suffered a psycholigical blow in childhood more often than chance would allow by making a specific prediction from it. The prediction is that the birth of the next youngest sibling occurred earlier in the lives of patients with cancer than in those of controls. The reasoning is basically simple. The birth of a younger sibling, with the consequent perceived loss of parental energy and time, is a traumatic event. Other things being equal, the earlier this occurs, the more severe the trauma. The child has had less time to be the baby and is younger and therefore less able to cope with his emotions; the younger the child, the more his feelings are on a preverbal and magical level and, therefore, more difficult to deal with.

METHOD

It was decided to choose a group of families which had been carefully evaluated from a medical standpoint, and which had at least one cancerous individual among the offspring. A set of pedigrees was chosen from Jacobsen's (1946) extensive data.[2] He studied 200 cancer patients, treated in Copenhagen hospitals, on whom very complete family records were available. Certain of Jacobsen's families were excluded. These were families in which the cancerous individual was the youngest sibling, a twin, or an only child, and those where the ages of some of the siblings were missing.

The controls were noncancerous offspring in the same families. It was believed that using these subjects would minimize genetic factors and make any systematic impact of other carcinogens in the data quite improbable.

As an additional test of the hypothesis, data on the *Ss* used in Reznikoff's study (1955) were reviewed with specific respect to the next youngest sibling. The *Ss* in this research were 50 women attending the Breast Clinic of Memorial Center for Cancer and Allied Diseases, New York, for the first time. It was eventually found that 25 of these women had malignant neoplasms, while the remaining half were subsequently diagnosed as having benign disorders. Also included in this study were 25 women from the Strang Detection Center declared to be entirely free of any breast pathology. The three groups were approximately the same in their age distributions. There were no group differences in average number of siblings, although the cancer *Ss* significantly less frequently had two or fewer sibs. The cancer and benign *Ss* were of comparable economic status, having been screened to insure that their incomes met the eligibility requirements of the Breast Clinic.

RESULTS

A chi square test of the frequencies in Table 1 shows that there is a statistically significant difference in duration of being the youngest child between the cancerous individuals and the controls. Individuals with cancer tend to have a shorter period of being the youngest child than do their cancer-free siblings. It is to be noted, however, that the difference falls almost entirely within the first interval of 10–23 months, which encompasses only 27% of the cancer cases and 9% of the controls.

As can be seen from Table 2, a significant difference was also obtained between Reznikoff's three groups, compared for age of next youngest sibling, using an analysis of variance design. From Table 3, which contrasts these three groups considered two at a time, it is further revealed that the cancer *Ss* were significantly closer in age to their next youngest sib than were the cancer-free individuals. Neither the comparison between the cancer and benign disorder groups nor between the benign disorder and cancer-free *Ss* resulted in significant differences, although the former approaches significance.

[1] This paper was written under the auspices of a grant from the Ayer Foundation, Inc.

[2] The writers wish to express their appreciation to the staff of the library of the American Cancer Society for their courtesy and help in selecting this set of pedigrees.

TABLE 1

FREQUENCY OF INDIVIDUALS WITH AND WITHOUT CANCER AS RELATED TO THE NUMBER OF MONTHS BETWEEN OWN BIRTH AND THAT OF THE NEXT YOUNGEST SIBLING[a]

	Months Between Own Birth and That of Next Youngest Sibling												
	10–23		24–35		36–47		48–57		58–64		64+		Total
	N	%	*N*	%	*N*	%	*N*	%	*N*	%	*N*	%	
Individuals with cancer	21	27	28	36	13	17	8	10	2	3	6	8	78
Cancer-free siblings	21	9	97	40	57	24	25	11	17	7	17	7	234

$\chi^2 = 18.32$, $df = 5$, $p < .01$.

[a] Jacobsen's families.

TABLE 2

COMPARISON OF CANCER, CANCER-FREE, AND BENIGN DISORDER GROUPS WITH RESPECT TO THE NUMBER OF MONTHS BETWEEN OWN BIRTH AND THAT OF THE NEXT YOUNGEST SIBLING

Source of Variation	*df*	Mean Square	*F*
Between groups	2	2558.51	3.78
Within groups	53	677.74	

$p < .05$.

TABLE 3

COMPARISON OF CANCER, CANCER-FREE, AND BENIGN DISORDER GROUPS CONSIDERED TWO AT A TIME WITH RESPECT TO THE NUMBER OF MONTHS BETWEEN OWN BIRTH AND THAT OF NEXT YOUNGEST SIBLING

Comparison	Mean difference months	*t*	*p*
Cancer with cancer-free	22.73	2.69	.02
Cancer with benign disorder	15.47	1.81	.10
Benign disorder with cancer-free	7.26	.85	.40

Note.—Means: cancer 26.53 months; cancer-free 49.26 months; and benign disorder 42.00 months.

DISCUSSION

A growing body of literature suggests that there are some correlations between personality organization and the occurrence of malignant disease. Two recent reviews (Kowal, 1955; LeShan, 1959) show that a large diversity of studies, using different methods, have come to this conclusion. Some studies seem to show indirectly that an early emotional trauma is found in a significant percentage of individuals who later develop cancer. Evans (1926) states that her cancer patients, who were in Jungian depth analysis, often regressed to a primary identification with the mother. In breast cancer patients, Bacon, Rennecker, and Cutler (1952) found a great deal of unresolved tension with the mother-figure. Cobb (1952) stated that her cancer patients regarded emotional relationships as dangerous and showed an absence of cathectic attachments. The papers of LeShan and Worthington (1956) and of Reznikoff (1955) point directly to an early traumatic event as being correlated with neoplastic disease.

It should be made clear that there is no implication here of a single cause of cancer. Our data suggest that under certain conditions, early emotional trauma may interact with other factors and play a part in the etiology of a later malignancy. The possible meanings of psychological events in the development of cancer have been partially explored by Meerloo (1954) and others, but much further research is needed to expand our gradually increasing awareness that psychological factors apparently must be viewed as being of some significance in the development of malignant disease, and to integrate this knowledge with the physiological findings in this field.

REFERENCES

BACON, C. L., RENNECKER, R., CUTLER, M. A psychosomatic survey of cancer of the breast. *Psychosomat. Med.*, 1952, **14**, 543–560.

COBB, B. A social-psychological study of the cancer patient. Unpublished doctoral dissertation. Univer. of Texas, 1952.

EVANS, E. *A psychological study of cancer*. New York: Dodd, Mead, 1926.

JACOBSEN, O. *Heredity in breast cancer*. London: Lewis, 1946.

KOWAL, S. Emotions as a cause of cancer. *Psychoanal. Rev.*, 1955, **42**, 217–227.

LESHAN, L. Psychological status as factors in the development of neoplastic disease: A critical review. *J. Nat. Cancer Inst.*, 1959, **22**, 1–18.

LESHAN, L., & WORTHINGTON, R. Some recurrent life history patterns observed in patients with malignant disease. *J. Nerv. Ment. Dis.*, 1956, **124**, 460–465.

MEERLOO, J. Psychological implications of malignant growth. *Brit. J. med. Psychol.*, 1954, **27**, 210–215.

REZNIKOFF, M. Psychological factors in breast cancer. *Psychosom. Med.*, 1955, **17**, 96–108.

(Received January 9, 1959)

Journal of Abnormal and Social Psychology
Vol. 60, No. 3, 1960

SOME RELATIONSHIPS BETWEEN GSR CONDITIONING AND JUDGMENTS OF ANXIETY[1]

HAROLD GILBERSTADT AND GENE DAVENPORT

University of Minnesota and Veterans Administration Hospital, Minneapolis

Previous studies have demonstrated differences in conditionability between samples of individuals grouped on the basis of anxiety. Classical conditioning procedures have been used, employing the eyelid reflex (Spence & Farber, 1953; Spence & Taylor, 1953; Taylor, 1951; Taylor & Spence, 1954) as well as the galvanic skin response (GSR) (Bitterman & Holtzman, 1952; Howe, 1958; Lykken, 1957; Welch & Kubis, 1957). The samples have included both normals and psychiatric patients. Anxiety has been assessed by self-inventories such as the Taylor Manifest Anxiety Scale (MAS) (Taylor, 1953) and by behavior rating scales such as that of Buss, Wiener, Durkee, & Baer (1955).

The mechanics of conditioning procedures have varied. However, two publications (Moeller, 1954; White & Schlosberg, 1952) have presented evidence that the probability of successful conditioning is increased by using a one-half second interval between the onset of the conditioned (CS) and unconditioned stimulus (US).

The aim of the present study was to investigate further the relationship between GSR conditioning and various clinical judgments of anxiety in a sample of psychiatric patients conditioned under presumably optimal procedures.

Method

Sample. Nineteen out of 23 consecutive admissions to the open wards of the Minneapolis VA Hospital constituted the sample. The four patients who were not used included two who showed no GSR to shock or buzzer and two who had to be eliminated because of apparatus irregularities. The sample included three depressive reactions, three psychophysiologic reactions, seven anxiety reactions, two emotionally unstable personality reactions, and one each of paranoid state, adult situational reaction, and incipient schizophrenic reaction. Mean Shipley-Hartford IQ was 112 on the Sines (1958) norms, with a range of 96 to 126. Mean age was 32 with a range of 21 to 57.

Apparatus. The apparatus consisted of a Fels Dermohmeter 22A and an Esterline-Angus recorder. A 2 cm. electrode was centered on each palm, and the discharge from a 2.0 microfarad condenser charged from a 75 volt source was applied across the third phalanges of the second and fourth fingers of the dominant hand.

Conditioning. Instructions were given which represented the investigation as a routine hospital test of physiological functioning. The conditioning procedure followed Lykken (1955, 1957) with the exception that a one-half second CS-US interval was used in place of the five-second interval. Adaptation consisted of ten trials in which a buzzer (No. 1) and a different-toned buzzer (No. 2) were presented intermittently but not followed by shock. Conditioning consisted of shock reinforced presentations of buzzer No. 1, interspersed with four nonreinforced presentations of buzzer No. 2. Extinction consisted of 24 presentations of the two buzzers without shock.

Measurement. The principal measure employed was an adaptation of the conditioning score devised by Lykken (1955) and consisted of the difference between (*a*) the ratio of the average GSR to buzzer No. 1 on the first three presentations of that buzzer during extinction, divided by the average GSR to shock on the last three presentations of shock during conditioning and (*b*) the ratio of the average GSR to buzzer No. 1 on the last three adaptation trials, divided by the GSR to shock on the first three conditioning trials.

Assessments of clinical anxiety. First, three clinical psychologists without knowledge of conditioning results independently categorized patients as high, medium, or low anxiety types by inspection of psychology folders containing admission notes and routine admission tests, including the MMPI. No projective tests were available to the judges. Second, the Buss (1955) rating procedure was used by the psychiatrist in charge of the case who rated his patient on the basis of his brief admission interview. Finally, patients were sorted into three approximately equal groups on the basis of the MAS items from the MMPI.

Results and Discussion

The relationships between the conditioning measure and the various anxiety measures are shown in Table 1. It can be seen that high, medium, and low anxiety groups as ranked by Clinician A, Clinician B, and the combined ratings of the three clinicians were significantly different in conditionability ($p < .05$) by the Kruskal-Wallis Analysis of Variance by Ranks test (Siegel, 1956). Neither

[1] We are indebted to J. J. Jenkins from whom we received helpful suggestions.

TABLE 1
RATED ANXIETY OF PATIENTS ORDERED ACCORDING TO CONDITIONING RANK

Item	Conditioning Rank																		Kruskal-Wallis Test	
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	H	P
Anxiety Criterion:																					
Clinician A	H	H	H	H	H	H	H	L	M	M	M	L	M	L	M	L	M	L	M	12.60	$P < .01$
Clinician B	H	H	M	M	H	H	L	L	H	M	M	M	M	L	M	L	M	L	M	6.34	$.02 < P < .05$
Clinician C	H	H	M	H	H	H	M	L	H	M	L	M	L	L	M	L	M	L	H	4.78	$.05 < P < .10$
Combined Clinicians	H	H	M	H	H	H	M	L	H	M	M	M	M	L	M	L	M	L	M	8.77	$.01 < P < .02$
MAS	M	H	L	H	M	H	[a]	H	H	M	L	M	M	L	H	L	M	L	H	2.91	$.20 < P < .30$
Psychiatric Rating	H	M	L	L	M	M	H	L	H	L	L	L	L	M	M	M	H	L	M	.17	$.70 < P$
Diagnosis:[b]	N	N	P	N	N	N	P	N	N	P	N	C	N	N	C	N	N	N	N		

Note.—H = High anxiety rating; M = Medium anxiety rating; L = Low anxiety rating.
[a] No MMPI taken by this patient.
[b] N = neurotic; P = psychotic; C = character disorder.

the rankings by the MAS nor the psychiatric ratings were significantly related to GSR conditioning.

Spearman rank correlation coefficients (Siegel, 1956) between basal resistance and the conditioning measure ($r = -.22$) and between basal resistance and the average shock response ($r = -.04$) were not significant, nor were there differences in extinction between the various anxiety groups.

It has been noted (Cronbach & Meehl, 1955) that while studying the validity of clinical constructs, it is important also to scrutinize the criterion measures. The results suggest that ratings of clinical anxiety by experienced clinicians based on psychological tests and history data (including the MMPI) covary with conditioning of the GSR. There is a suggestion in the present investigation, as in other recent investigations (Goldberg, 1959), that studies employing judges should have some basis for inferring the expertness of the judges with the particular instruments. In the present instance, Clinician A, who was presumed to be most competent and most experienced with the MMPI was apparently most successful, and Clinician C, who was presumed to be least competent and least experienced with the MMPI, was apparently least successful (See Table 1).

Present findings are consistent with those of Bitterman and Holtzman (1952), in which judgments of anxiety by clinicians with several types of information available were significantly related to GSR conditioning, whereas the MAS was not.

The psychiatric ratings based on clinical observation from brief admission interviews were not significantly related to conditioning in the present instance, although in the Welch and Kubis study (1947) clear-cut relationships were obtained, possibly because the psychiatrists had greater familiarity with the patients.

One factor which could in part account for the lack of significant differences in extinction between groups was revealed during posttest interviews. Some patients stated that during the long series of buzzers in the extinction trials they "expected" a few shock stimulations. Cook and Harris (1937) have shown that such a factor can override the effects of conditioning and extinction.

SUMMARY

Using presumably optimum GSR conditioning procedures with a one-half second US-CS interval and an ongoing sample of 19 psychiatric patients rated for clinical anxiety, (*a*) anxiety groups ranked on the basis of admission psychological data by clinical psychologists were found to be significantly different in conditionability; (*b*) anxiety groups ranked either on the basis of the Taylor Manifest Anxiety Scale or brief psychiatric admission interviews were apparently not significantly different on conditioning; and (*c*) anxiety groups revealed no significant differences in extinction.

REFERENCES

BITTERMAN, M. E., & HOLTZMAN, W. H. Conditioning and extinction as a function of anxiety. *J. abnorm. soc. Psychol.*, 1952, **47**, 615–623.

BUSS, A. H., WIENER, M., DURKEE, ANN, & BAER, M. The measurement of anxiety in clinical situations. *J. consult. Psychol.*, 1955, **19**, 125–129.

COOK, S. W., & HARRIS, R. E. Verbal conditioning of the galvanic skin reflex. *J. exp. Psychol.*, 1937, **21**, 202–210.

CRONBACH, L. J., & MEEHL, P. E. Construct validity in psychological tests. *Psychol. Bull.*, 1955, **52**, 281–302.

GOLDBERG, L. R. The effectiveness of clinicians' judg-

ments: the diagnosis of organic brain damage from Bender-Gestalt test. *J. consult. Psychol.*, 1959, **23**, 25–33.

HOWE, E. S. GSR conditioning in anxiety states, normals, and chronic functional psychotic subjects. *J. abnorm. soc. Psychol.*, 1958, **56**, 183–189.

LYKKEN, D. T. A study of anxiety in the sociopathic personality. Unpublished doctoral dissertation, Univer. Minnesota, 1955.

LYKKEN, D. T. A study of anxiety in the sociopathic personality. *J. abnorm. soc. Psychol.*, 1957, **55**, 6–10.

MOELLER, G. The CS-UCS interval in GSR conditioning. *J. exp. Psychol.*, 1954, **48**, 162–166.

SIEGEL, S. *Nonparametric statistics*. New York: McGraw-Hill, 1956.

SINES, L. K. Intelligence test correlates of Shipley-Hartford performance. *J. clin. Psychol.*, 1958, **14**, 399–404.

SPENCE, K. W., & FARBER, I. E. Conditioning and extinction as a function of anxiety. *J. exp. Psychol.*, 1953, **45**, 116–119.

SPENCE, K. W., & TAYLOR, JANET A. The relation of conditioned response strength to anxiety in normal, neurotic, and psychotic subjects. *J. exp. Psychol.*, 1953, **45**, 265–272.

TAYLOR, JANET A. The relationship of anxiety to the conditioned eyelid response. *J. exp. Psychol.*, 1951, **41**, 81–92.

TAYLOR, JANET A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285–290.

TAYLOR, JANET A., & SPENCE, K. W. Conditioning level in the behavior disorders. *J. abnorm. soc. Psychol.*, 1954, **49**, 497–503.

WELCH, L., & KUBIS, J. Conditioned PGR (psychogalvanic response) in states of pathological anxiety. *J. nerv. ment. Dis.*, 1947, **105**, 372–381.

WHITE, CARROLL T., & SCHLOSBERG, H. Degree of conditioning of the GSR as a function of the period of delay. *J. exp. Psychol.*, 1952, **43**, 357–362.

(Received October 13, 1958)

Journal of Abnormal and Social Psychology
Vol. 60, No. 3, 1960

VERBAL REINFORCEMENT IN SCHIZOPHRENIA[1]

EDWIN COHEN
Veterans Administration Hospital, Iowa City, Iowa

AND BERTRAM D. COHEN[2]
Lafayette Clinic, Detroit

THE common clinical observation of impaired "reality contact" in schizophrenia implies a deficit in the intensity or direction of responsiveness to social cues. Social cues, and specifically those involved in verbal interactions, have been shown to provide reinforcement effects in a wide variety of experimental settings (Salzinger, in press). If schizophrenic patients manifest faulty responsiveness to social cues, it may be predicted that the reinforcing function of listener reactions (i.e., verbal reinforcement) will be negligible for them as compared with nonschizophrenic control *S*s. The present experiment was designed to test this prediction. The particular technique used was one with which reinforcement susceptibility had been previously demonstrated for general medical patients (Cohen, Kalish, Thurston, & Cohen, 1954) and for psychiatric patients selected on the basis of anxiety level scores rather than diagnosis (Taffel, 1955).

METHOD

Apparatus. The materials consisted of 160 3×5 white index cards, on each of which was typed a common verb (used by Cohen et al., 1954). All verbs were in the past tense. Six pronouns—I, We, He, They, She, and You—were placed below each verb. The order of pronouns was randomized for the series of 160 cards.

Subjects. *S*s were 32 male hospitalized patients from the psychiatric service of the VA Hospital in Iowa City. All were capable of reading and free of CNS pathology. By unanimous agreement of the staff—two psychiatrists and two clinical psychologists—half of the *S*s had been diagnosed as schizophrenic, the other half as neurotic. Both groups were unselected as to subtype of illness and were comparable in age and education. Mean age of the schizophrenics was 32, while that of the neurotics was 32.4; mean education was 11.3 and 10.5 years for the schizophrenics and neurotics

[1] Portions of the present paper were read at the Midwestern Psychological Association meetings in St. Louis, Missouri, May, 1956.

[2] At VA Hospital, Iowa City, Iowa, when study was conducted.

respectively. The groups varied significantly in the number of days they had been hospitalized; schizophrenics mean time was 130 days, with a range of 10 to 523, whereas the neurotics were hospitalized for a mean of 57 days with a range of 18 to 166 days.

Procedure. Each *S* was informed that he was participating in a research project. He was then shown a sample card and instructed to make up a sentence containing the verb on the card and beginning with any one of the pronouns. *S* then constructed sentences in this fashion until the 160 cards were completed. *E* recorded each sentence verbatim. The cards were shuffled for each *S* to randomize the effect of order.

The 16 *S*s in each diagnostic category were randomly divided into two equal groups. One group in each diagnostic category was reinforced for any sentence begun with I or We, and the other group for similar use of He or They. During the acquisition phase (the first 80 cards), *E* said "Good" in a flat, unemotional tone following any sentence which *S* began with the pronouns chosen to be reinforced. In order to determine a base level without reinforcement, however, reinforcement was not administered until after the first 20 card presentations.

The last 80 presentations constituted the extinction phase of the experiment. During this phase, reinforcement was discontinued for all *S*s. In order to test *S*'s awareness of the relation between his responses and those of *E*, each *S* was asked at the completion of the series which pronoun he had used most, as well as specific questions about his ideas concerning the purpose of the experiment and *E*'s responses.

RESULTS

In the treatment of the data, the number of reinforced pronouns occurring within 8 successive blocks of 20 cards each were used as the criterion measures, taking the first four blocks as the acquisition and the last four as the extinction phases. A summary of the analysis of the acquisition trials is shown in Table 1. Neither Methods nor any of its interactions with other factors approaches reliable significance. It may therefore be safely assumed that the *S*s conditioned equally to both classes of pronouns, I-We and He-They. The *F* values for Trials and Trials-by-Diagnosis are each significant at the 2% level. It seems, therefore, that conditioning did occur, but it occurred differentially for the two diagnostic groups. Since it appeared from inspection of Fig. 1 that the curve for the schizophrenic *S*s was not ascending, the data for this group alone were further analyzed. The two methods of conditioning were combined, since the initial analysis had shown no differences attributable to methods. The *F* value for Trials was clearly not significant, allowing the inference that no reliable conditioning occurred for the schizophrenic group. It may then be concluded, because the over-all Trials effect for all groups is significant, that conditioning occurred in the neurotic group.

Since the neurotic group had demonstrated acquisition, their data for Trial Blocks 4 through 8 were analyzed for extinction. As suggested in Fig. 1, no reliably significant *F* was obtained for Trials. No evidence of extinction was found.

Initial questioning of *S*s with regard to the purpose of the experiment and the possibility that *E* had influenced their responses in any way re-

TABLE 1
SUMMARY OF ANALYSIS OF ACQUISITION DATA

Source of Variation	*df*	*ms*	*F*
Between Subjects	31		
Diagnosis	1	82.89	
Methods	1	2.82	
Diagnosis × Methods	1	.38	
Error (b)	28	34.89	
Within Subjects	96		
Trials	3	32.15	3.96*
Trials × Diagnosis	3	28.57	3.52*
Trials × Method	3	5.38	
Trials × Diagnosis × Method	3	7.70	
Error (w)	84	8.12	
Total	127		

* Significant 2% level.

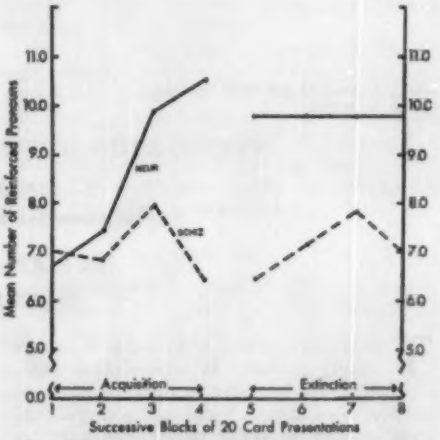

FIG. 1. Mean numbers of reinforced pronouns for successive blocks of 20 card presentations.

vealed no clear evidence of awareness of the reinforcement contingency. Further questioning of a more specific nature, in which *E* explicitly asked if his saying "Good" had any effect on *S*'s responses, did result in "insight" statements by one schizophrenic and five neurotics. The acquisition scores for these *S*s inicated no better conditioning than the scores for the remaining 26 who furnished no evidence of awareness.

DISCUSSION

The acquisition results confirm the hypothesis that reinforcement effects are generally negligible for schizophrenic patients. This held true for both response classes to which the reinforcer was applied (either first or third person pronouns). In contrast, the neurotic groups showed significant acquisition curves. Again, the response class made no difference.

The combined acquisition curve for the two neurotic groups resembles that obtained for a group of nonpsychiatric patients reported in a previous paper (Cohen et al., 1954). The combined acquisition curve for the schizophrenic groups resembles that obtained for the nonreinforcement control group in the previous study. In other words, saying "Good" to the schizophrenic *S*s was apparently functionally equivalent to saying nothing to nonschizophrenic *S*s.

The failure to demonstrate extinction as a consequence of nonreinforcement, as seen in the results for the schizophrenic group, is to have been expected since no acquisition was demonstrated. On the other hand, acquisition was demonstrated by the neurotics, and yet extinction was not achieved under nonreinforcement. This finding is in accord with data for nonpsychiatric patients under a comparable extinction procedure (Cohen et al., 1954). The possible explanation cited in that study, referring to the existence of continuing sources of reinforcement inherent in the procedure, also applies to this study. Another possible explanation involves the relatively low frequency of the reinforced response at the initiation of the extinction procedure. It may be noted from Fig. 1 that the frequency of I and We for the neurotic *S*s was less than 11 for 20 cards at the time extinction was initiated. Since a previous study (Cohen et al., 1954) continued the reinforcement procedure for another 80 cards without any increase in frequency of the reinforced pronoun class, the solution to increasing its frequency would not seem to lie in additional reinforcement trials prior to initiation of extinction procedures. One possible solution may be to decrease the number of possible different responses, thus raising the initial level and the final acquisition level as well. For example, You and She could be eliminated and I, We, He, and They retained.

Certain limits on the generality of the present results, regarding the nonsusceptibility to reinforcement of schizophrenic patients, should be specified. These include the limits set by (*a*) properties of the stimulus class employed as a reinforcer and (*b*) properties of the response class subjected to reinforcement operations.

In the present experiment, the reinforcing stimulus (Good) was one that connotes social approval. *E*'s explicit responses were administered so as to provide reward for the desired response class rather than punishment of responses in a competing class. Results summarized by Rodnick and Garmezy (1957) suggest that schizophrenic *S*s are especially sensitive to verbal punishment. It is possible that the aversive properties of verbal punishment could result in more normal acquisition curves for schizophrenic *S*s. A recent study by Leventhal (1958) provides data in support of this prediction.

Studies of motor learning (Cohen, 1956), reaction time (Rosenbaum, Mackavey, & Grisell, 1957), and operant behavior (Lindsley, 1956) of schizophrenic *S*s support the hypothesis of Hunt and Cofer (1944) that psychological deficit in schizophrenia reflects a loss of social motivation based on a process of extinction of responsiveness to acquired incentives and rewards. If schizophrenic patients have not incurred some irreversible impairment of the ability to respond to acquired reinforcers, one might predict that the experimental pairing of a stimulus like the word Good with a primary reinforcer would result in its acquisition (or reaquisition?) of reinforcing properties.

In an interview-type setting, Salzinger and Pisoni (in press) applied verbal reinforcing stimuli (Recognition and Agreement) to schizophrenic patients' verbal responses of a "personal-affective" type. Significant operant acquisition effects were demonstrated. No changes were found in nonaffective self-reference responses during the same interview. In the present experimental procedure, the reinforcer was applied to responses in the prearranged classes, regardless of any specific affective meanings the responses may have had for the individual patients.

A study by Hartman (1955) suggests that the initial (operant level) response subjected to reinforcement operations may be of particular importance for schizophrenic patients. Using a technique similar to that of the present study, Hartman subdivided his *S*s in accordance with the initial levels of preference they showed for the to-be-reinforced pronoun class. For his schizophrenic *S*s, he found significant acquisition effects for the

subgroup which showed a high initial preference for the reinforced pronoun class. In the present study, reinforcement was administered to responses without regard for prereinforcement preferences.

Impaired susceptibility to verbal reinforcement, in view of the present results, warrants consideration as an objective, experimentally manipulable correlate of schizophrenic behavior. It is recognized that many investigators conceive of schizophrenic impairment of responsiveness to social cues (e.g., disturbances in reality contact) in terms of changes in perceptual and cognitive variables. Studies can be designed to clarify the relations of these variables to those of verbal reinforcement. In any event, further systematic investigation of the various parameters of verbal reinforcement in normal and pathological groups appears both warranted and promising.

REFERENCES

COHEN, B. D. Motivation and performance in schizophrenia. *J. abnorm. soc. Psychol.*, 1956, **52**, 186–190.

COHEN, B. D., KALISH, H. I., THURSTON, J. R., & COHEN, E. Experimental manipulation of verbal behavior. *J. exp. Psychol.*, 1954, **47**, 106–110.

HARTMAN, C. H. Verbal behavior of schizophrenic and normal subjects as a function of types of social reinforcement. Unpublished doctoral dissertation, State Univer. Iowa, 1955.

HUNT, J. McV., & COFER, C. N. Psychological deficit. In J. McV. Hunt (Ed.) *Personality and the behavior disorders*. New York: Ronald, 1944. Pp. 971–1032.

LEVENTHAL, A. The effects of differential verbal reinforcement on psychiatric and nonpsychiatric hospital patients. Unpublished doctoral dissertation, State Univer. Iowa, 1958.

LINDSLEY, O. R. Operant conditioning methods applied to research in chronic schizophrenia. *Psychiat. res. Rep.*, 1956, No. 5, 118–139.

RODNICK, E., & GARMEZY, N. An experimental approach to the study of motivation in schizophrenia. In M. R. Jones, (Ed.) *Nebraska symposium on motivation*. Lincoln: Univer. Nebraska Press, 1957. Pp. 109–184.

ROSENBAUM, G., MACKAVEY, W. R., & GRISELL, J. L. Effects of biological and social motivation in schizophrenic reaction time. *J. abnorm. soc. Psychol.*, 1957, **53**, 364–368.

SALZINGER, K. Experimental manipulation of verbal behavior: a review. *J. gen. Psychol.*, in press.

SALZINGER, K., & PISONI, STEPHANIE. Reinforcement of affect responses of schizophrenics during the clinical interview. *J. abnorm. soc. Psychol.*, in press.

TAFFEL, C. Anxiety and the conditioning of verbal behavior. *J. abnorm. soc. Psychol.*, 1955, **51**, 496–501.

(Received April 2, 1959)

Journal of Abnormal and Social Psychology
Vol. 60, No. 3, 1960

EGO DISJUNCTION:

A FAILURE TO REPLICATE TREHUB'S RESULTS

WILBERT E. FORDYCE
Veterans Administration Hospital, Seattle, Washington

AND WILLIAM R. CROW
University of Washington

TREHUB (in press) introduced the concept of ego disjunction (ED), defining it as "the concurrence in an individual of two or more relatively strong, relatively long enduring needs or characterological dispositions with mutually incompatible objectives." ED was measured by the Edwards Personal Preference Schedule (EPPS) Edwards, 1954), on which seven pairs of a prior incompatible needs were taken to indicate the level of ED. Trehub found a monotonic linear relationship between ED scores on the EPPS and diagnostic category, ranging from low ED for college students, up through adolescents, neurotics, character disorders, and schizophrenics. The last three groups were patients at a VA Neuropsychiatric hospital.

This study replicated that part of Trehub's study involving patient groups. In addition to the ED measure, a series of quantitative indices of disorganization based on the MMPI were compared with ED scores and with the diagnostic criterion. MMPI measures used (all expressed in *T* scores) were Pt-Sc plus 100 (Delta); Sc plus 2 (Sc-Pt), with Sc scores above 90 considered as 90

TABLE 1
MEANS AND STANDARD DEVIATIONS OF DIAGNOSTIC GROUPS ON ED AND SELECTED MMPI SCORES

Groups		N	Scores				Social Desirability
			Ego Disjunction	Delta	Gamma	Triad	
Neurotic	Mean	22	52.59	98.91	73.68	202.68	26.86
	SD		23.82	7.02	17.69	36.24	7.38
Character Disorder	Mean	31	50.42	100.16	68.06	193.48	27.94
	SD		22.93	11.66	30.16	34.59	6.65
Acute Schizophrenia	Mean	20	51.85	96.60	84.90	223.35	23.05
	SD		20.61	8.14	22.82	38.17	7.50
Chronic Schizophrenia	Mean	31	59.48	86.94	101.19	218.94	25.03
	SD		28.57	13.44	34.95	43.89	7.58

(Gamma); Pa plus Pt plus Sc (Triad); and the 39-item form of the Edwards (1953) Social Desirability scale (SD).[1]

Initially the diagnostic groups studied were drawn from a population of adult, male, veteran, neuropsychiatric patients at the VA hospital (GM and S), Seattle, Washington. There were 20 acute schizophrenics (Ac-Sc), 31 character disorders (Ch-Dis), and 22 neurotics (Neu) as classified by a single, final diagnosis assigned by the hospital staff. No cases with multiple psychiatric diagnoses were used. Later, a further sampling of more chronic schizophrenic patients was made to extend the range of severity of disorganization tapped by the tests. Under the same criterion of final hospital diagnosis, chronic male schizophrenics (Ch-Sc) were added—21 from Western State Hospital, Steilacoom, Washington, and 10 from the VA hospital (NP), American Lake, Washington.[2] These two subsamples were combined.

All subjects performed on the EPPS and the booklet form of the MMPI either the same day or within a very few days of each other.

RESULTS

The data failed to duplicate Trehub's findings.

Table 1 shows the means and standard deviations on ED and the MMPI scores for each diagnostic group. Table 2 shows the *t* test comparisons of the means, indicating no differences between any of the criterion groups on the ED measure. In contrast, the MMPI indices of disorganization more consistently discriminate the schizophrenic groups from the character disorder group, though not from the neurotic group.

DISCUSSION

The failure of the ED measure to discriminate schizophrenic patients from the character disorder or neurotic patient groups raises a serious question as to the validity of the EPPS measure of the ego disjunction construct. While no adolescent or college student groups were included in this study, the range of presumed levels of ego disjunction represented in the four patient groups studied should be considerable. The success of the MMPI indices in discriminating the groups lends support to the inference that a range of sickness levels or ego disjunction does exist in the groups studied There was no criterion contamination in the chronic schizophrenic group because the MMPI was administered after diagnoses were made. In the other groups, the MMPI as a part of the diagnostic workup process may have contributed in varying amounts to final diagnosis. Criterion contamination could account for differences found between neurotic and character disorder groups and the acute schizophrenic group. It is obvious, however, that far more than the MMPI was involved in arriving at a final diagnosis.

Failure of ED to discriminate when the MMPI indices succeeded suggests that sampling differences may have influenced Trehub's original findings. Differences in diagnostic criteria may have entered into these conflicting findings. It seems possible that cultural differences in test behavior between patients from New England and from the Pacific Northwest may contribute to the different findings of the two studies.

The fair success of the MMPI measures used here in discriminating psychotics from nonpsychotics offers some promise as a quantitative, objective criterion for use in other studies.

[1] The Delta and Gamma scores were defined arbitrarily and are designed to treat differently the shape and elevation factors of the Pt and Sc scales.

[2] We are indebted to F. E. Shovlain, Superintendent; R. W. Brown, Clinical Director; and Mack F. Knutsen, Supervising Psychologist, all of Western State Hospital; and to Leon S. Diamond, Chief of Professional Services and John Marks, Chief, Psychiatric Research Unit, VA hospital, American Lake, Washington, for their help in providing data for this study.

TABLE 2
DISCRIMINATION OF DIAGNOSTIC GROUPS BY *t* TESTS ON ED AND MMPI INDICES

Groups	Scores								*SD*	
	ED		Delta		Gamma		Triad			
	t	*p*	*t*	*p*	*t*	*p*	*t*	*p*	*t*	*p*
Neu–Ch-Dis	.384	.80	.440	.70[a]	.769	.50[a]	.899	.40	.545	.60
Neu–Ac-Sc	.104	.90	.962	.40	1.475	.10	1.756	.10	1.614	.20
Neu–Ch-Sc	.907	.40	3.750	.01[a]	2.650	.02[a]	1.399	.20	.855	.40
Ch-Dis–Ac-Sc	.222	.90	1.167	.30	2.092	.05	2.834	.01	1.224	.30
Ch-Dis–Ch-Sc	1.354	.20	4.080	.01	3.935	.01	2.496	.02	1.582	.20
Ac-Sc–Ch-Sc	1.013	.40	2.833	.02[a]	1.810	.10[a]	.361	.80	.896	.40

[a] Correction for unequal variances made by calculating *df* as $N - 1$ of the smaller group.

SUMMARY

Trehub's EEP-derived ego disjunction measure was compared with a series of MMPI-derived measures of ego disorganization on groups classified by psychiatric diagnosis. In contrast to Trehub's original study, the ego disjunction measure failed to discriminate among neurotics, character disorders, acute schizophrenics, and chronic schizophrenics. All MMPI measures except the Social Desirability scale discriminated schizophrenics from nonschizophrenics.

It was concluded that sampling differences, differences in diagnostic practice, or cultural differences in patient groups studied could account for the conflicting findings of this study and the original work by Trehub.

Success of the MMPI indices in discriminating psychotics from nonpsychotics offers promise of the utility of these indices in other studies.

REFERENCES

EDWARDS, A. L. *Experimental design in psychological research.* New York: Rinehart, 1950.

EDWARDS, A. L. *Personal preference schedule.* New York: Psychological Corp., 1954.

EDWARDS, A. L. The relationship between the judged desirability of a trait and the probability that the trait will be endorsed. *J. appl. Psychol.*, 1953, **37**, 90–93.

TREHUB, A. Ego disjunction and psychopathology. *J. abnorm. soc. Psychol.*, in press.

(Received April 20, 1959)

Journal of Abnormal and Social Psychology
Vol. 60, No. 3, 1960

A NOTE CONCERNING HOMOGENEITY OF MEMBERSHIP AND GROUP PROBLEM SOLVING

MARVIN E. SHAW

Massachusetts Institute of Technology

HOMOGENEITY of group members with regard to individual differences has long been recognized as an important variable in group behavior (Hemphill, 1950). In a recent article, Hoffman (1959) reports results supporting the hypothesis that homogeneity results in less efficient problem solving than does nonhomogeneity of membership. Citing the findings of Pelz (1956) to support his view, he suggests that "... the results reported in the present study are probably generalizable well beyond the limited population of college students who supplied the data." (Hoffman, 1958, p. 31). It was this statement which motivated the writer to report some of his own observations of relations between homogeneity and group effectiveness.[1]

During the past few years, the writer has conducted a series of experiments attempting to study the joint effects of group structure and

[1] The purpose of this note is not to criticize Hoffman's excellent report, but only to offer a word of caution regarding the generalizability of the results and a word of encouragement to others who may be interested in this problem.

TABLE 1

THE RELATION OF MEMBER HOMOGENEITY TO GROUP PERFORMANCE AND SATISFACTION

Homogeneity Variables	Structural Conditions	N	Group Process Variables		
			Trials to Target	Time Scores	Ratings of Satisfaction
Intelligence (Shaw, 1959a)	Power:				
	Centralized	20	+.07	+.38	−.49*
	Decentralized	20	−.07	+.08	+.21
	Communication:				
	Centralized	24	—	.00	+.21
	Decentralized	24	—	+.26	+.30
Acceptance of Authority (Shaw, 1959a)	Power:				
	Centralized	20	−.11	−.17	−.08
	Decentralized	20	+.48*	+.53**	−.59**
	Communication:				
	Centralized	24	—	.00	+.02
	Decentralized	24	—	+.27	−.11
Individual Prominence (Shaw, 1959b)	Power:				
	Centralized	22	−.22	+.06	−.18
	Decentralized	22	−.08	−.26	+.20

Note.—Signs of correlations have been adjusted so that a positive correlation indicates increasing efficiency or satisfaction with increasing homogeneity.

* $p < .05$.

** $p < .01$.

composition upon group efficiency in problem solving situations. (Shaw, 1959a, 1959b). While not related to the primary purpose of the investigations, the data were analyzed systematically to determine the relationship between various estimates of homogeneity and problem solving effectiveness under varying structural conditions. The results of these analyses are given in Table 1. The measure of homogeneity used in computing the correlations reported was the average deviation of scores earned by members of the group on measures of intelligence (the verbal scale of the Scholastic Aptitude Tests), acceptance of authority, and individual prominence. Essentially the same results were obtained, however, when other indices of homogeneity were used, such as range and standard deviation. Measures of group process variables were number of trials to criterion, total time required by the group to complete the assigned task, and average ratings of satisfaction by group members.

A number of comments may be made in connection with the results reported in Table 1. First, there seems to be no systematic relation between homogeneity and group efficiency and satisfaction. Degree and direction of correlations vary with the group structure and the particular characteristic under consideration.

Second, most of the correlations fail to meet the usual requirements for statistical significance. This may be due to the small number of cases in each sample, to the fact that the homogeneity estimate is based upon a single characteristic rather than a profile as were Hoffman's measures, or to the fact that no relation exists. It is interesting to note, however, that the one condition yielding uniformly significant correlations (acceptance of authority in a decentralized power structure) is precisely the one where such a relation might be expected on a priori grounds.

Third, centralized structures appear consistently to produce correlations that differ from those found in decentralized structures, although these differences are mostly unreliable.

These findings point to the difficulties in generalizing the results of a single study or even several studies beyond the limits imposed by the experimental situations. It may be, of course, that the effects of profile homogeneity will prove to be unchanged by such extraneous variables as group structure, group task, and the like; however, it seems far more likely that the effects of homogeneity, like so many other variables in group behavior, will be found to vary with the particular conditions under which its effects are measured.

The general problem of homogeneity as a variable in group behavior is a tremendously important one and deserves extensive investigation. The limited evidence reported here, along with the findings of Hoffman (1959) and Pelz (1956), indicates that the results of such studies will richly repay the investigator for his efforts.

REFERENCES

HEMPHILL, J. K. The measurement of group dimensions. *J. Psychol.*, 1950, **29**, 325–342.

HOFFMAN, L. R. Homogeneity of member personality and its effect on group problem-solving. *J. abnorm. soc. Psychol.*, 1959, **58**, 27–32.

PELZ, D. C. Some social factors related to performance in a research organization. *Administrative Sci. Quart.*, 1956, **1**, 310–325.

SHAW, M. E. Acceptance of authority, group structure and the effectiveness of small groups. *J. Pers.*, 1959, **27**, 196–210. (a)

SHAW, M. E. Some effects of individually prominent behavior upon group effectiveness and member satisfaction. *J. abnorm. soc. Psychol.*, 1959, in press. (b)

(Received March 4, 1959)

THE BODY IMAGE BARRIER CONCEPT[1]

CHARLES R. EIGENBRODE AND WILLIAM G. SHIPMAN

Staunton Clinic, School of Medicine, University of Pittsburgh

A GOOD deal of the recent research on the body image has been done by Fisher and Cleveland. Their initial study of rheumatoid arthritics (Cleveland & Fisher, 1954) led them to focus on the external limits of the body image, which may serve unconsciously as a defensive barrier to ward off threats from the outside and to hold disturbing, hostile impulses within. A scoring system for Rorschach protocols was derived which yields a Barrier score—an index of the degree of concern a subject has about his body limits. This measure was validated on a group of patients with internal and external disorders (Fisher & Cleveland, 1955), then applied with interesting findings to cancer patients (Fisher & Cleveland, 1956b), to style of life (Fisher & Cleveland, 1956a), and to physiological reactivity (Fisher & Cleveland, 1958; Fisher, 1959). It has been assumed in these papers that the measure represents a psychological factor that is not only a strong determinant of locus of disease but is related to central personality tendencies. The purpose of the present study was to determine how well the Barrier score differentiates between patients with internal and external psychosomatic illnesses.

METHOD

Sample. The Rorschach records used in this study were obtained from the case files of Staunton Clinic. The medical record of each patient was checked so that assignment to one of the experimental groups could be made. No patient was included in the sample who (*a*) had symptoms of both internal and external disease, (*b*) was below 16 or above 50 years of age, or (*c*) had less than 10 responses in his Rorschach. Cases were drawn so that the two patient groups were similar in age, education, and IQ. Two groups of records were obtained: 54 cases of psychosomatic skin disorders (26 atopic eczema, 16 chronic urticaria, and 12 miscellaneous skin disorders) and 29 cases of internal psychosomatic disorders (18 stomach ulcer, 8 ulcerative colitis, and 3 genitourinary disease). Emotional factors were considered by the physicians to be the major factor in all skin cases.

Scoring. After learning the scoring system, the authors collaboratively scored the 83 records, which had been randomly mixed and from which all primary identification material had been removed. The number of scored responses was equated for the two groups by setting a maximum of 20 scored responses per record. This control was successful in that the median number of responses per record in both groups was 20. After three to five days, 30 of the 83 records were pulled at random and rescored as a check on the consistency with which the scoring system had been applied. This comparison yielded a Pearson correlation of .95. At this point, Fisher was consulted. His scoring of 10 of these records yielded a Pearson r of .84 with that of the authors. This consultation indicated that several basic changes in the scoring system used by the authors needed to be made in order to coincide more closely with Fisher and Cleveland's actual practice. This prompted a rescoring of the 83 records by the two authors individually. The extent of their agreement was indicated by a Pearson r of .80, which compares favorably with an r of .82 reported in the original monograph (Fisher & Cleveland, 1955).

[1] The authors wish to thank Seymour Fisher and Stanley King for personally clarifying the scoring system. Our associate, Leslie Berger, made many helpful suggestions in the latter stages of the project.

RESULTS AND DISCUSSION

A nonparametric median test was applied to the Barrier score data of both of the writers for the interval vs. the external disease groups. Neither of the obtained chi squares approached significance. Indeed, the obtained frequencies were almost identical with those expected by chance. In the light of this absence of difference in the Barrier scores obtained by patients with internal and external diseases, it seems doubtful that a concern with firmness of body boundaries plays a significant role in determining the location of an illness in the body.

Where does the reason for this discrepant finding lie: in the sample, in the measuring instrument, in the theory, or in some combination?

The sample differs from that used in the original study mainly in disease distribution and sex ratio. In this study, the sex ratio for the various diseases was kept similar to the known distribution of the disorders in the general population. But sex differences should not be an important source of discrepancy since both Fisher and Cleveland and the present investigators found no sex difference in the Barrier score.

The scoring system may have been misapplied or have intrinsic weaknesses that allow different results in different studies. The correlation of .84 between the initial scoring of the writers and Fisher, plus the subsequent rescoring, suggests that reliable scoring was not a problem. This turns the focus on the Barrier score itself, which seems open to considerable criticism. The scoring rules, *in detail*, have not been published. No retest reliability has been reported. And most important of all, no cross-validation of the scale seems to have been published by other investigators. The small size of the modal Barrier score, 4, and the range, 0 to 12 in this study, allows great instability of scores. Thus, a subject's classification as "high" or "low" may also be unstable.

This highlights another variable, the examiner. There is a great range in the amount of inquiry that is done during the administration of a Rorschach, and the presence or absence of additional descriptive material most certainly affects the opportunity for Barrier scores. To check on Barrier score differences between examiners, a nonparametric analysis of variance by ranks was done on the scores obtained by the seven examiners who tested 78% of the cases used here. The resultant Kruskal-Wallis H was statistically nonsignificant, and it was concluded, therefore, that differences between examiners did not cause significant differences in the number of Barrier responses given by the subjects in this sample.

Conceivably, the scale does not measure the Barrier concept very well. In many of the major scoring categories there is a great deal of apparent arbitrariness about which Rorschach responses fitting the category may be scored. For example, when considering animals with unusual skins, a mink is scored but a possum is not. Other inconsistencies emerge in the "unusual skin" category which may contribute a great deal to the total. And why is not the opposite end of the dimension scored? One would think that minus scores for such soft surfaced things as sponges and wool would improve the scoring. Neglecting to score responses at one end of the continuum should hypothetically greatly reduce the sensitivity of the measure. These various points lead one to conclude that there are some striking gaps in the application of the Barrier principle to the scoring system.

The theory itself warrants careful inspection. The scoring instructions refer to "an individual's perception of his body boundaries as *definite* and as possessing *protective* properties." There are obviously two features involved here, and it is a question whether these two features, definiteness and protectiveness, are necessarily related. A person may be well aware of his limits but not necessarily fearful of attack and defensively oriented. Presumably, body limits are clearer with emotional maturity, whereas a preoccupation with defense to external threat occurs at a less mature level. Similarly, the Barrier theory scores with equal weight, *containers* and *defensive covering*. Though these both involve a hard, encircling surface, the points of view involved seem quite different. One wonders if these two concepts really represent only one psychological variable or whether there are at least two factors involved. The point, then, is that the Barrier concept may entail different and not entirely consistent dimensions. This complexity of the over-all measure would be acceptable if the role of the subfactors were known and the measure correlated successfully with other variables. But under the present circumstances, some of the subcategories may be measuring the Barrier tendency while others, which do not, cancel the relation when the subcategories are combined.

SUMMARY

A check was made on whether the Rorschach Barrier scores of 54 patients with external psychosomatic disorders were greater than those of 29 patients with internal psychosomatic disorders. Statistical analysis showed no evidence of difference in the distribution of Barrier scores in the two groups. The potentialities of the Barrier approach cannot be truly assessed until ambiguities in the theory are clarified and until the scale is improved as a measuring instrument.

REFERENCES

CLEVELAND, S. E., & FISHER, S. Behavior and unconscious fantasies of patients with rheumatoid arthritis. *Psychosom. Med.*, 1954, **16**, 327–333.

FISHER, S. Prediction of body exterior vs. body interior reactivity from a body image schema. *J. Pers.*, 1959, **27**, 56–62.

FISHER, S., & CLEVELAND, S. E. The role of body image in psychosomatic symptom choice. *Psychol. Monogr.*, 1955, **69** (17, Whole No. 402).

FISHER, S., & CLEVELAND, S. E. Body image boundaries and style of life. *J. abnorm. soc. Psychol.*, 1956, **52**, 373–379. (a)

FISHER, S., & CLEVELAND, S. E. Relationship of body image to site of cancer. *Psychosom. Med.*, 1956, **18**, 304–309. (b)

FISHER, S., & CLEVELAND, S. E. *Body image and personality*. Princeton, N. J.: Van Nostrand, 1958.

(Received August 26, 1958)

Neurology and Psychiatry

Section VIII of Excerpta Medica

The Neurology Section includes abstracts of all articles on neurological surgery in addition to general clinical neurology and experimental neurophysiology.

The section on Psychiatry covers the topics usually contained within the meaning of the word. In addition the following special branches are dealt with: Adlerian psychology, individual and analytical psychology; sexology, criminology, alcoholism, and drug addiction as related to psychiatry; psychosomatic medicine; social and industrial psychology and psychiatry, vocational guidance and mental testing; psychoanalysis; personality testing and the Rorschach test; heredity and statistical studies as they apply to psychiatry. The subjects of mental defect and epilepsy are fully reported.

Annual volume of 1000 pp., including an index classified by author and subject

Subscription price $22.50 per annual volume

EXCERPTA MEDICA FOUNDATION

The New York Academy of Medicine Building

2 East 103 Street New York 29, N.Y.

RESEARCH IN PSYCHOTHERAPY

Proceedings of an interdisciplinary conference, Washington, D. C. April 9-12, 1958
Edited by A. Rosenzweig and Morris B. Parloff, Editors

Presentations and discussions by leading investigators in psychotherapy research, both psychologists and psychiatrists, combine to produce a critical evaluation of research objectives and methodologies.

Prepared papers by J. C. Whitehorn, J. D. Frank, Lewis Robbins and R. S. Wallerstein, Timothy Leary and Merton Gill, C. R. Rogers, George Saslow and J. D. Matarazzo, J. I. Lacey, E. S. Bordin, and W. U. Snyder present comprehensive descriptions of large-scale research programs and extensive analyses of crucial research problems.

Formal discussion papers by J. M. Butler, David Shakow, Milton Greenblatt, and Maurice Lorr, and informal discussion by all the participants provide expert commentary on present scopes of research. The summary chapter by the editors is an orderly presentation of the major issues covered by the conference.

Price $3.00

Order from:

AMERICAN PSYCHOLOGICAL ASSOCIATION
Department RP
1333 Sixteenth Street, N.W.
Washington 6, D.C.